Assessing the role of the Benguela Current as a major biogeographic barrier to marine coastal fishes.

Michael Gwilliam

A thesis submitted to Aberystwyth University in partial fulfilment for the degree of Doctor of Philosophy.

Institute of Biological, Environmental and Rural Sciences
Aberystwyth University

2016

Declaration

Word Count of thesis:	55,923
This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree.	
Candidate name	MICHAEL GWILLIAM
Signature:	
Date	25/12/2016

STATEMENT 1

This thesis is the result of my own investigations, except where otherwise stated. Where *__correction services__ have been used, the extent and nature of the correction is clearly marked in a footnote(s).

Other sources are acknowledged by footnotes giving explicit references. A bibliography is appended.

Signature:	
Date	25/12/2016

[*this refers to the extent to which the text has been corrected by others]

STATEMENT 2

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signature:	
Date	25/12/2016

Acknowledgements

I would like to thank first and foremost my Supervisor Professor Paul Shaw and Dr Niall McKeown. They have both been of exceptional help in writing this thesis in what at times were not ideal circumstances for me. Additionally they have ensured I haven't gone too far off the academic rails over the last few years.

I would also like to thank the help of my colleagues in the Shaw Lab namely Dr Max Blake ~~(soon to be Dr)~~ and Amy Rainbow Unicorn Moonchild Healey for all their help and discussions during the last few years. In particular I would like to acknowledge their help over the last couple of days to ensure submission of this thesis!

Our (retired) lab assistant Alun Evans has also been of great help in ensuring the smooth operation of the laboratory work and a great source of advice in where to find and get things done in Aber! I would also like to thank my colleague Dr Hayley Watson for all her help and advice, particularly for initially showing me and Max how to genotype microsatellites (admittedly after being bribed with a bottle of rosé).

My Colleagues in South Africa Dr Warren Potts and Alex Winkler have been integral to this project and to whom (along with their colleagues) I would have none of my southern African samples. I would also like to recognise Dr Olgac Guven for collecting my Turkey outgroup samples which have proved invaluable in subsequent analyses. Many thanks also to Sean Roche and the 'Celtic Spirit' for attaining our Aberystwyth samples of Black Seabream.

My parents have also been of immense help the last couple of years and somewhat integral in ensuring I kept with the PhD to get it submitted! My Friends and family have also been of great help in keeping me motivated.

Many Thanks to you all!

Dedicated to my late friend Torin Lakeman.

Abstract

Contemporary biodiversity has been shaped by interactions between intrinsic characters of organisms and extrinsic environmental features that often cannot be reliably predicted. Regions where many species boundaries coincide represent intriguing systems to disentangle factors underpinning eco-evolutionary divergence with relevance to the understanding of how species may respond to climate change. Such regions often harbour complex patterns of inter- and intra-specific biodiversity that are taxonomically challenging. The Benguela cold upwelling system (BCS) has emerged as a major biogeographic barrier promoting and maintaining differences between the Atlantic Ocean and Indo-Pacific Ocean coastal fauna. This research assessed nuclear and mitochondrial DNA variation and applying population genetic and phylogeographic analytical frameworks assessed divergence across the BCS in a number of coastal Sea Bream species (Sparidae). In conjunction with available morphological and ecological data, the results support species level divergence across the BCS in what were previously regarded as conspecific populations in both *Spondyliosoma* and *Diplodus cervinus / hottentotus* and direct taxonomic rearrangement. The data also support the potential inadequacy of mtDNA COI barcoding for species discovery in the region due to variable interspecific coalescent depths. *Lithognathus mormyrus* and *Sarpa salpa* revealed similar phylogeographic patterns with (1) distinct African and Atlantic / Mediterranean and (2) divergence across the BCS. Phylogenetic analysis supported the ancestral status of African clades and the potential of an Angolan glacial refuge. A salient feature of the results was the signature of historical and largely asymmetric gene flow across the BCS from South Africa to Angola. This episodic permeability has implications for unexpected species' responses to climate change which are discussed. Collectively, this study provides new insights into the dynamic role of the BCS in shaping marine biodiversity in both the Atlantic and Indo-Pacific regions.

Contents

Chapter 1: Introduction

1.1 Introduction

In the terrestrial environment population structuring has been traditionally attributed to the combined effects of distance, geographical barriers to movement / dispersal and life history traits (Holderegger and Wagner, 2008). The same processes, albeit on a larger and longer scale, are also attributed to the evolution of new species (i.e. speciation events) in the terrestrial environment (Coyne and Orr, 2004). In marine systems the proposed mechanisms for population sub-structuring and speciation are less well understood, mainly due to the relative inaccessibility of the marine environment. In the past the initial presumption has been of no population structuring within marine species due to the largely homogenous environment the oceans present, presumed ease of dispersal by adults or early life stages (larvae) through the oceans, and (historically) large population sizes (Palumbi, 1992). With advances in molecular biology techniques it has been revealed over the last twenty years that there is significant population sub-structuring within many marine species, as well as cryptic speciation events. The present study aims to investigate how environmental features and changes in these features interact with species' biology and life history traits, driving evolutionary processes in African coastal fishes. To achieve these aims the project primarily utilises molecular population genetic and phylogenetic techniques.

1.2 Molecular genetic markers

Genetic markers are normally sequences of DNA used to identify individuals or species. An ideal genetic marker needs to be variable (polymorphic) to enable identification, and typically markers with slower mutation rates are utilised to identify higher order taxa (e.g. species, using mtDNA COI sequences) whilst genetic markers such as microsatellites, with a faster mutation rate are used to identify individuals within a population and 'shallow' population structuring. Genetic markers can also be either selectively neutral or adaptive, whereby any mutation within the genetic marker sequence which infers an advantage / disadvantage in terms of the organisms overall fecundity is termed a selective or adaptive marker (Holderegger *et al.*, 2006). Neutral genetic markers are thus useful for establishing the evolutionary history of the organism

in question due to neutral processes such as gene flow, demographic changes and stochastic environmental events, whilst non-neutral markers may be useful for establishing traits driving adaptive population subdivision and speciation (Holderegger *et al.*, 2006).

1.2.1 Microsatellites

The use of microsatellite DNA markers in population genetics research has intensified over the last two decades. Due to their properties of having a short sequence length and consequent ease of amplifying during PCR, and their highly polymorphic nature, they are excellent candidate markers for determining population structure (Chistiakov *et al.*, 2006; Selkoe and Toonen, 2006). Microsatellites are DNA regions (or loci) that comprise a tandem repeated sequence of 1 – 5 bases. For example $(AC)_{20}$ is a microsatellite that repeats the sequence AC twenty times. The repeat sequence is flanked by 'unique' (i.e. non-repeated) DNA sequence within which primer sequences for PCR can be sited (Estoup and Cornuet 1999). There can be up to a million microsatellite loci within the genome of a species; this along with their tandem repeat signature makes them more easily identifiable for producing molecular markers. The high mutation rate within microsatellites is suggested to be due to mistakes during DNA replication (thought to occur as much as every 1000 replications; Ellegren, 2004). These mistakes or 'slips' result in the addition or subtraction of repeat units, resulting in predictable and stable changes in the size of the PCR amplified fragment that can be used to identify different sequence alleles (Estoup and Cornuet, 1999).

Whilst microsatellites have many favourable attributes and are widely used in molecular ecology studies they do come with some distinct disadvantages. Firstly whilst ubiquitous they can be costly and difficult to initially develop, whilst some taxa exhibit high failure rates during development e.g. marine invertebrates (e.g. Cruz *et al*., 2005), lepidopterans (Meglecz *et al*., 2004). Secondly mutational mechanisms are unclear in microsatellites (elaborated below). Thirdly, their high mutation rates may result in homoplasy, e.g. two alleles may be identical in sequence but not identical by descent (i.e. they have differing genealogical histories). Homoplasy is expected to be most extensive when mutation rate is high but this can be accounted for in analyses (Slatkin, 1995; Estoup and Cornuet, 1999). Finally null alleles can plague microsatellite studies (Selkoe and Toonen, 2006). Null alleles may arise from mutations in the flanking

priming sites resulting in failure of the primer to anneal during PCR, whilst poorly preserved DNA may result in a bias to shorter repeat alleles.

1.2.2 Microsatellite models of evolution

To estimate differentiation between two sets of microsatellite data theoretical models of microsatellite mutation are needed (Balloux and Lugon-Moulin, 2002). There are three major models of evolution for microsatellites: stepwise, infinite alleles and two phase models. The Stepwise Mutation Model (SMM) assumes that each mutation in the microsatellite DNA sequence involves the addition or deletion of a single repeat unit, and so alleles are related by descent in a predictable way (Kimura and Ohta, 1978). The Infinite Alleles Model (IAM) assumes that mutations involve random changes in the number of repeat units, and so alleles are related by descent in no predictable way (Kimura and Crow, 1964). The Two Phase Model (TPM) is a mixture of the SMM and the IAM, in that most mutations involve a single stepwise change in length but a limited proportion of mutations involve changes of multiple repeats (Valdès *et al*., 1993; Di Rienzo *et al*., 1994).

1.2.3 mtDNA

Mitochondrial DNA (mtDNA) is a circular DNA molecule found in mitochondria. The mitochondrial genome is made up of 13 protein coding genes, 22 transfer RNAs and two ribosomal RNAs. There is also a Control Region containing sites for replication and transcription initiation. Most sequences in the mitochondrial genome are non-repetitive, with few spacer sequences between genes or intervening sequences within transcribed genes. Mitochondrial genomes can be presumed haploid; they experience no recombination, and are maternally inherited.

mtDNA markers are the most prevalent markers used to infer historical and evolutionary relationships for genetic lineages (Galtier *et al.*, 2009). mtDNA markers have several advantages that help make them so popular. First, the mitochondrial genome has a conserved arrangement of sequences; this means many universal primers for PCR amplification are available. Second, mtDNA has higher rates of mutation accumulation than most coding nuclear DNA regions, making mtDNA more likely to exhibit sequence variation. Third, mtDNA is maternally inherited, giving it appealing dynamics for studying sex-biased population processes and hybridisation between differing populations. Fourth, there is no recombination, making it easier to trace

genetic lineages and so reconstruct historical processes in the evolution of the population or species. Fifth, mtDNA exhibits a small effective population size (mtDNA is haploid and only inherited from one parent, so has an effective population size that is one quarter that of the nuclear genome), and so is more susceptible to the effects of genetic drift and more readily displays the effects and signals of demographic changes in populations (such as past bottlenecks).

With these advantages there are however some distinct disadvantages to mtDNA. Firstly, mtDNA acts effectively as a single locus and therefore there is no scope for comparing genealogies of multiple independent loci. Secondly, the small effective population size may exaggerate historical events and underestimate genetic diversity. Thirdly, it only provides a genetic picture of female-linked population processes in the evolution of the species. Fourthly, the selective neutrality of mitochondrial genes has been called into question with some studies suggesting the mitochondrial genes could also be involved in speciation (Galtier *et al*., 2009). Finally, mtDNA sequences have been found in nuclear DNA, known as pseudo-mtDNA or nu-mtDNA. If the mtDNA is present as a pseudo gene in the nuclear genome it will not be subject to many of the above characteristics of mtDNA and will evolve differently and thus may confuse analysis and interpretation of data thought to be mtDNA in origin – see Galtier *et al.* (2009) for a critical review of mtDNA.

1.3 Population structure of marine organisms

Marine species have long been considered to be panmictic, that is to say that individuals within a species potentially are free to move throughout the species range and reproduce with any other individual (Palumbi, 1992). Panmixia in marine species was presumed to be supported by two common features: that they have very large ranges, and that many are free to disperse throughout their range either as adults or during their larval stages. However research over the last two decades has to some extent refuted this notion. Studies using modern genetic techniques have found that marine species may exhibit complex population structures, ranging from complete panmixia to extensive population divergence resulting in speciation (Sala-Bozano *et al.*, 2009; Henriques, 2012; Winkelmann *et al*., 2013; Bowen *et al*., 2016; Thompson *et al*., 2016).

1.3.1 Landscape genetics and phylogeography

Landscape genetics studies how features in the landscape and environmental variables influence allele frequencies among populations (Holderegger and Wagner, 2008). A population living in a homogeneous landscape would be expected to show panmixia. Likewise a population living in a landscape with many structural features such as mountains, rivers, etc., might be expected to show more structuring as these features may present barriers to gene flow. Landscape genetics has much crossover with metapopulation concepts. Metapopulations can best be described as populations within a population, with each population being either more or less successful (a source, in terms of migrants) or unsuccessful (a sink, for migrants), often associated with favourable or unfavourable 'niche' conditions (Hanski, 1999). A strict metapopulation model would involve each population having equal chances of extinction and of being recolonised from successful populations: this is however a rather simplistic and unrealistic model. More sophisticated concepts involve source populations that are stable through time, which continually re-colonise sink populations (Hanski, 1999). Genetically sink populations may be recognised by having low genetic diversity, whilst source populations may exhibit high genetic diversity. The amount of contemporary gene flow is also of importance to identify in a sink population: if gene flow is high, then present genetic diversity may also be artificially high. Such a scenario is of importance since it could lead to establishing conservation areas where inherently a species would not thrive if the source population was threatened. Phylogeography aims to understand how evolutionary relationships among genetic lineages and their geographical locations influence current distributions of genetic diversity (Avise, 2000; Avise *et al*., 2016). Phylogeography thus attempts to understand how historical (evolutionary) and spatial (geographical distribution) factors interplay to determine current observed distributions of species and genetic lineages within species.

Landscape genetics has a similar goal to phylogeography but is only concerned with how present day geographical and environmental factors determine a population's location and structure, whilst phylogeography focuses not only on present day factors but historical environmental, geographical and evolutionary processes. More recently there has been movement for an amalgamation of the two approaches under the broader term of biogeography (Rissler, 2016). Landscape genetics studies often use highly polymorphic genetic data from markers such as microsatellites, SNPs (Single

Nucleotide Polymorphisms) and AFLPs (Amplified Fragment Length Polymorphisms) as well as environmental and spatial data (Storfer *et al*., 2010). Geographical Information Systems (GIS) are used to map the genetic data and its correlation with environmental factors such as temperature, primary productivity, annual rainfall and seasonality. Phylogeography often uses mtDNA sequences to infer evolutionary history: however since mtDNA acts as a single locus effects such as hybridization or selection on the locus could skew resultant relationships. Therefore more recently mtDNA has been compared for concordance of evolutionary history to nuclear sequences and plastid DNA (in plants and microbes; Avise, 2009; Karl *et al*., 2012).

1.3.2 Applying terrestrial population genetics concepts to the sea

Both phylogeography and landscape genetics were developed almost exclusively for studying terrestrial and freshwater systems. However both approaches can and are being applied to marine systems. The broadly termed 'Seascape genetics' is based upon terrestrial landscape genetics theory and practice (Selkoe *et al*., 2008). Therefore seascape genetics can be broadly defined as the influence of seascape features and marine environmental variables on gene frequencies within marine species populations. Although the term seascape genetics is becoming increasingly popular in the literature, it should be noted that it is by no means ubiquitous - many published studies that could be classified as seascape genetics do not use the term. Phylogeography has however been successfully applied to marine systems, with many studies in phylogeography focusing on marine species (e.g. Duncan *et al*., 2006; Kelly *et al*., 2006; Lessios, *et al*., 2003). Before proceeding it is worth noting that often studies incorporate both 'seascape genetics' elements and phylogeography, hence in this section the studies mentioned are not divided as such.

The distribution of genetic diversity in marine species is influenced by barriers to dispersal (Palumbi, 1992). Population divergence in marine species is, however, often comparatively weak compared to terrestrial species (Waples, 1998). This is thought to be due to the ease of dispersal in the marine environment compared to terrestrial systems, as both aquatic adults and pelagic larvae can more easily move or float with prevailing currents over large distances (Galarza *et al*., 2009). Additionally marine taxa have (at least historically) large population sizes and display greater effective population sizes (N_e) compared to terrestrial taxa, as such marine taxa have a lower level of genetic drift reducing observed genetic population structuring (Palumbi, 1992).

There is however clear evidence that marine species can exhibit complex and deep genetic structuring (Kelly and Palumbi, 2010), often associated with physical barriers to contemporary dispersal. Historical events such as glacial periods, changing currents and habitats may also influence present day structuring of marine organisms (Roy *et al*., 1995; Wares, 2002). The relative importance of present day versus historical factors in determining abundance and distribution of marine taxa remains unclear, and a major focus of marine population genetics.

There are well-described biogeographic boundaries in the marine realm varying in age, intensity and permeability to dispersal and gene flow across them by marine species. For example the Antarctic Circumpolar Current formed 20-30 million years ago (Ma) isolating the coastal Antarctic fauna (Kock, 1992), whilst the uplift of the Isthmus of Panama 3 Ma separated the Pacific Ocean and Atlantic Ocean marine fauna (Avise, 2000). Historical sea level changes during glacial cycles resulted in fragmented lineages of marine taxa in Southeast Asia, where the extensive Sunda shelf was exposed during glacial periods (aka the Marine Wallace Line; Barber *et al*., 2000; Lourie and Vincent 2004). Geological (plate tectonics) and historical glacial cycles have also promoted the observed phylogenetic breaks in marine taxa around the Los Angeles and Monterey Bay area along the North American coast (previously attributed to Point Conception; Burton, 1998; Dawson, 2001). More recently occurring boundaries include examples such as the North Sea - Baltic Sea divergence established ~7,500 years ago (Olsen *et al*., 2004), and the Northern Atlantic shelf boundary (Cape Hatteras, east coast of North America) where the Labrador and Gulf Stream currents meet (Weinberg *et al*., 2003). These more recent biogeographic boundaries usually were established and subsequently influenced by climatic fluctuations (glacial cycles) during the Quaternary period.

Barriers to species dispersal (and so gene flow) in the marine realm can be referred to as being ‘hard’, such as the formation of land barriers which physically split ocean regions and their constituent populations, or ‘soft’, which involve hydrographical processes (e.g. currents, fronts, large distances) which disrupt movements of marine organisms either as adults or during larval stages. A recent review of vicariance across major marine biogeographic boundaries found both hard and soft barriers to be of comparable ‘evolutionary strength’ in terms of temporal concordance and intensity (Cowman and Bellwood, 2013).

Physical boundaries are perhaps the most obvious 'hard' barriers to dispersal in marine systems. The uplift of the Isthmus of Panama 3 Ma has led to clear divergence of taxa to either side of the land barrier, for example in sea urchins and snapping shrimps (Hurt *et al*., 2009; Lessios *et al*., 2003; McCartney *et al*., 2000; Taylor and Hellberg, 2005). Similarly, the outflow from the mouth of the Amazon River presents a physical barrier of fresh water to along-coast dispersal by marine coastal fauna (Joyeux *et al*., 2001). These major barriers to dispersal are often associated with allopatric speciation.

More recently, oceanographic features such as currents, upwelling's, fronts and eddies have been found to have strong influences in shaping population structure and even evolutionary history of marine species. Nikaido *et al.* (2011) studied the genetic population structure of Coelacanth (*Latimeria chalumnae*) populations in southern Africa and northern Tanzania using mtDNA. The northern Tanzanian Coelacanth population was genetically divergent from the southern populations, which Nikaido *et al.* (2011) postulate is due to diverging ocean currents: the north flowing East African coastal current and the south flowing Mozambique current present a barrier to gene flow between the northern and southern populations.

1.3.3 Case Study: Mediterranean Sea

The Mediterranean Sea fauna have been closely examined from a phylogeographic perspective. Furthermore all the species in the present study have ranges which extend into the Mediterranean. As such it is prudent to outline the phylogeography of this region. There are over 8500 macroscopic species identified as native to the Mediterranean, representing between 4% and 18% of the global marine biodiversity, with more than a quarter of these species being endemic (Longhurst, 1998; Bianchi and Morri, 2000). This is remarkably high diversity and endemism for a semi-enclosed basin representing only 0.82% of the surface area and 0.32% of the volume of world oceans (Patarnello *et al*., 2007). This high diversity may be due to the geological history of the Mediterranean, the most extreme event being its complete desiccation during the Messinian Salinity Crisis (MSC) 5.33 Ma, which lasted half a million years. This period of desiccation drove the pre-existing Indo-Pacific biota to extinction, being replaced by an East Atlantic-derived biota during the re-flooding of the Mediterranean through the Straits of Gibraltar (Almada *et al*., 2001; Domingues *et al*., 2005). Following the MSC, from 3 Ma onwards the Mediterranean underwent cold glacial and warmer interglacial

periods until the present interglacial. This geological history suggests that the present day Mediterranean biota consists of (Patarnello *et al*., 2007):

I. Temperate Atlantic species.
II. Endemic species (Miocenic) and neo-endemic species (Pliocenic).
III. Subtropical Atlantic species (interglacial remnants).
IV. Boreal Atlantic species (glacial remnants)
V. Red Sea invasive (Lessepsian) species entering through the Suez Canal.
VI. Recently invasive North East Atlantic species.

Patarnello *et al.* (2007) suggest three geographical provinces in the Mediterranean: the eastern Atlantic, western Mediterranean and eastern Mediterranean. Some species show panmixia across these regions whilst others show complete genetic isolation between the Mediterranean and Atlantic areas, with the Almeria-Oran frontal system being identified as the actual location of the Atlantic-Mediterranean boundary.

Genetic studies have illustrated the complex population structures, and the suggested underlying processes, that may occur within species across a hydrologically and ecologically complex area such as the Mediterranean Sea. For example, Sala-Bozano *et al*. (2009) found that present day population structure (indicated by microsatellite markers) in the Seabream *Lithognathus mormyrus* showed a major break between the East Atlantic and the Mediterranean Sea across the Almeria-Oran Front, with further population sub-structuring within the Mediterranean Sea. mtDNA analyses indicated that there was also deeper genetic divergence between the Atlantic and Mediterranean populations of *L. mormyrus*, suggesting that the two populations were previously isolated for 3 Myrs due to geological transformations before and during the Pleistocene, but regardless of the long period of time since divergence individuals from each clade freely interbreed upon secondary contact. Pérez-Losada *et al.* (2007) also explored the boundary between the Atlantic and Mediterranean Sea in cuttlefish (*Sepia officinalis*) and found isolation-by-distance to be the dominant factor determining population structure in the region, with some evidence highlighting historical isolation due to segregation of water bodies during the Pleistocene. Zardoya *et al*. (2004) found contrasting population genetic structures in the Mediterranean Sea, and where the chub mackerel (*Scomber japonicus*) exhibited a largely panmictic population with extensive gene flow whereas the mackerel (*Scomber scombrus*) exhibited genetic differentiation and structuring along an east-west axis. Zardoya *et al*. (2004) concluded that these

observed differences in population structure were due to either lower dispersal capabilities in *S. scombrus* or to differing life history strategies. These studies highlight the intricacies and complexities of the factors underpinning population structure in marine systems and the need for a combination of genetics, oceanography, palaeoceanography and biology to more fully understand population structure.

1.3.4 Life history strategies

Life history strategies are how organisms schedule key characteristics throughout their lifetime to maximise fecundity; as such these traits should be fine-tuned by natural selection. Important life history characters include: size and number of offspring; age at sexual maturity; reproductive lifespan; maximum body size. These life history characteristics are often interrelated with one another: for example large body size is related with a later onset of sexual maturity and increased size of offspring in mammals (Western, 1979).

Understanding how life history characters may promote or inhibit population structuring and speciation events in the oceans is a lively area of research. Broadly speaking several life history traits are of particular interest: egg type (either pelagic or direct developing (i.e. retained by the parent)), inshore or offshore spawning, pelagic larval duration, adult body size, coastal or oceanic dwelling adults, and habitat preference of adults (e.g. pelagic, demersal or benthic). Each of these traits has the potential to determine a species ability to disperse, thereby increasing or decreasing potential gene flow (Galarza *et al*., 2009)**.**

Of all the above traits the length of pelagic larval duration (PLD) has long been presumed the best correlate to dispersal, due to the number of marine species exhibiting this trait and the ease of transport of larvae by oceanic currents. However as indicated above many fish species exhibit highly structured and complex genetic populations, suggesting that many pelagic larval stages fail to realise their dispersal potential. Overall the evidence for the importance of early life history characters in determining genetic population structure is mixed. For example Shulman and Bermingham (1995) found early life history strategy to be a poor predictor for genetic population structure in Caribbean reef fishes, whilst Purcell *et al*. (2006) found early life history strategy to be a good predictor of population genetic structure in the reef fish *Haemulon flavolineatum*. Similarly two studies focusing on fish populations in and around the Gulf of California found contrasting results: Riginos and Victor (2001) found early life

history strategy was a good predictor for observed genetic population structure in three blennioid reef fish species, whilst Bernardi *et al*. (2003) concluded that early life history traits (specifically PLD) were a poor explanation of the observed genetic differences. Galarza *et al*. (2009) studied seven littoral fish species representing six different families, and how oceanic fronts in the Western Mediterranean and early life history strategies relate to population structure. Again Galarza *et al*. (2009) found early life history traits not to be a good predictor for population structure; they concluded that further work was warranted to study adult life history traits and how oceanic features can disrupt gene flow regardless of a long PLD. One additional problem is highlighted by the Bernardi *et al.* (2003) study: they could not find early life history data for all of the species they studied, quite simply because it has not been studied.

1.3.5 Life history characteristics: Sex change

In contrast to most vertebrates, hermaphroditism is particularly common in fish, with many being sequential hermaphrodites (i.e. maturing as one sex and then changing sex later in life; Warner, 1988). All the species in the present study are sequential hermaphrodites. Hermaphroditism has arisen independently in 23 teleost families (Frisch, 2004). Ghiselin (1969) proposed three models (or advantages) for the evolution of hermaphroditism in animals. Firstly the low density model postulates that hermaphroditism increases reproductive success since individuals have 100% chance of meeting suitable mates (in true hermaphrodites) compared to 50% in gonochoristic (non sex changing) species, plus true hermaphrodites can also potentially self fertilise in the absence of available mates. The second hypothesis is the size advantage model; this model is perhaps the simplest explanation and the most obvious advantage for sequential hermaphroditism. Simply put, reproductive success must be increased in smaller (younger) individuals as one sex and then would be greater as the opposite sex in larger (older) individuals. This hypothesis can also explain why protandry and protogyny may evolve. Generally in ectotherms, females reproductive success is positively correlated to body size (Warner, 1975), since eggs are larger than sperm, so older larger females can produce more eggs and have increased reproductive success; such a scenario thus highly promotes protandry. Conversely protogyny may evolve where there is male territoriality, competition for mates and male brooding, and since larger individuals can better fight for females and defend their territories / nests. A third and final explanation for the evolution of hermaphroditism according to Ghiselin (1969)

is the ‘gene dispersal model’, which is based upon the limitations of dispersal ability and its subsequent effects upon genetic population structure. There are two versions of this model: the inbreeding version and the sampling error version. The inbreeding version of the model postulates that in small populations protandry / protogyny prevent inbreeding amongst siblings, which can then only breed with the older (and more likely genetically distinct) individuals in the population. The sampling error version of the model recognises that the genetic environment is prone to phenomena such as genetic drift (and similar phenomena), especially in small populations which would bias the sex ratio, so sequential hermaphroditism increases reproductive success (and population genetic diversity).

Contrary to general assumptions sequentially hermaphroditic species sex ratios are skewed towards the first sex to reach sexual maturity. Garratt (1985) studied the South African protogynous Sparid *Chrysoblephus puniceus*, finding that sex ratios were highly biased to females (up to 90% in some catches). In a wider study Allsop and West (2004) analysed 121 sex changing species (including 76 fishes) finding protogynous species to be heavily biased to females and protandrous species being heavily biased to males. Such skewed sex ratios are well known to reduce the effective population size (Hartl and Clark, 1997). As such it can be expected that this would lead to an increased level of genetic drift in sequential hermaphrodites leading (in turn) to an increase in genetic population structuring. Chopelet *et al*. (2009) in a meta-analysis studied 99 species of marine fish, 19 of which were protogynous and seven were protandrous, with the overall hypothesis that hermaphroditic fish would exhibit more genetic population structuring than their gonochoristic counterparts. Overall the study failed to identify any such pattern after controlling for confounding variables such as dispersal ability, egg type, sample size, maximum body size and habitat use. Chopelet *et al*. (2009) are cautious in interpreting the result since even though confounding variables were controlled in the study there was still potential that any effect of sex change on population structuring was masked by the reduced power of the model to test for the effects of sex ratio. Regardless of any reduced power to detect the effects of sex ratio in the study its relative importance must be marginal in comparison to other confounding variables such as dispersal ability, egg type, territoriality, etc. Whilst it is logical that sex bias could be resultant from hermaphroditism it is also known that individual reproductive success can cause a discrepancy between the census size and effective size

in many marine taxa particularly in species that are highly fecund and can have significant infant and juvenile mortality (Hauser and Carvalho, 2008). Whilst at the intra-species level hermaphroditism seems to have a limited effect on population genetic structuring it still could be highly important for individual species and is deserving of further evaluation in this study.

1.3.6 Speciation in marine systems

Research into speciation over the last fifty years has focussed on (and largely still does) the prevalence of species evolving allopatrically (i.e. in geographical isolation) or sympatrically (within the same geographical location). For a new species to evolve a population needs to be reproductively isolated (i.e. no gene flow). In allopatric speciation a population is divided by a barrier, through which individuals cannot disperse (i.e. no gene flow); these isolated populations may then undergo independent adaptation to their respective environment and become genetically and morphologically distinct. When and if the two diverged populations meet again they cannot reproduce with one another due to having undergone independent evolution to the point of distinct species. Simple isolation of a population, however, may be insufficient to result in speciation.

Although allopatric speciation has been considered the main means new species evolve it is now recognised that new species can evolve due to the influence of environmental gradients within the species range (parapatric speciation), within isolated populations at the edge of a species range (peripatric speciation), or in sympatry (Galarza *et al*., 2009). Selection pressures by the environment (such as temperature and food sources) have been found to be a strong factor promoting speciation (Coyne and Orr 2004; Schluter, 2001; Schluter and Conte, 2009), and a number of instances have been described of sympatric speciation (e.g. Dieckmann and Doebeli, 1999; Elmer and Meyer, 2010; Savolainen *et al*., 2006). Speciation also can arise by hybridisation forming 'instant species', not only in plants but in animals and microbes (see Mallet (2007) for a review). Establishing whether a new species resulted from allopatric, parapatric, peripatric or sympatric speciation is often hard, often due to shifts in species' ranges through time (Barraclough and Vogler, 2000; Quenouille *et al*., 2011).

1.3.7 Speciation in the sea

Physical barriers to dispersal and gene flow such as the closure of the Isthmus of Panama, the establishment of the Benguela Current upwelling system (see below), the freshwater outflow from the Amazon River and low availability of food in the open ocean all have potential to cause significant population structuring and speciation in marine organisms.

'Hard' and 'Soft' barriers - allopatric speciation

Establishing whether oceanographic features ('soft' barriers) such as the Benguela Current system and solid land barriers ('hard' barriers) such as the Isthmus of Panama have similar potential as isolation barriers is of primary importance to our understanding of how new species can evolve in marine systems. Similarly does the open ocean present a reproductive barrier to wide ranging coastal species? For example, Claremont *et al*. (2011) studied the marine snail *Stramonita haemastoma* species complex to establish whether the Atlantic Ocean presents a dispersal barrier to these coastal species, and thus acts as a speciation mechanism. Claremont *et al*. (2011) found that despite the long PLD of *Stramonita*, speciation has occurred within the Atlantic, both in response to barriers operating at the largest geographical scale (the width of Atlantic) and at smaller scales within the western Atlantic. This finding suggests that the open ocean can present a substantial barrier, comparable to land bridges such as the Isthmus of Panama, which has also caused speciation in *Stramonita* (Claremont *et al*., 2011).

The role of genetic drift and gene flow in marine speciation

The levels of gene flow among populations within species, and the consequent effects of genetic drift, are strong factors promoting speciation (Coyne and Orr, 2004). Marine systems may provide excellent examples of how these two factors interplay during the speciation process since fully impermeable barriers to gene flow are rare in marine systems, albeit with the exception of land barriers. If full genetic isolation is seldom reached then genetic drift may play a lesser role in speciation in marine systems, with genetic drift being further impeded by the larger N_e in marine species. The role of selection and adaptation in speciation (in contrast to genetic drift) in marine systems could be important but can be hard to quantify due to problems such as difficulty in sampling, lack of expertise and data on morphology and lack of detailed knowledge of

the general ecology of marine compared to terrestrial organisms. Chiba *et al*. (2009) found Sparidae jaw morphology to be an important convergent evolutionary trait: this could suggest a role of ecological selection since jaw morphology is highly adapted to what a fish can eat. However, genetic drift has been found to play a major part in structuring of marine fish species populations through bottlenecks and founder events where N_e is reduced (Henriques, 2011; Von der Heyden *et al*., 2007).

1.4 The study system: the Benguela Current

On geological timescales the Atlantic and Indo-Pacific Oceans were inter-connected with marine coastal taxa able to freely disperse globally between oceans. However, from around 34 Ma onwards resulting from interactions of plate tectonic processes and associated changes in global oceanographic systems, several major biogeographic barriers developed that isolated the Indo-Pacific and Atlantic coastal faunas. The first barrier arose as a result of the onset of the Antarctic Circumpolar Current from 34 Ma, which presented a cold water barrier to warm temperate coastal fauna migrating around the southern tip of South America (Kock, 1992) along with the extirpation of any warm temperate coastal taxa found in Antarctic waters (DeConto and Pollard, 2003). The second major barrier to arise was the closure of the Tethys seaway around 14 Ma which presented a physical land barrier to marine fauna dispersal between the northern Atlantic and Indian Oceans. Finally the uplift of the Isthmus of Panama presented another clear physical land barrier between the central West Atlantic and eastern Pacific oceans (Lessios, 2008). After the closing of tropical-temperate connections between the oceans the only remaining route for connection of coastal fauna between the Indo-Pacific and Atlantic oceans was in southern Africa around Cape Agulhas and through the Benguela Current system (Figure 1.1).

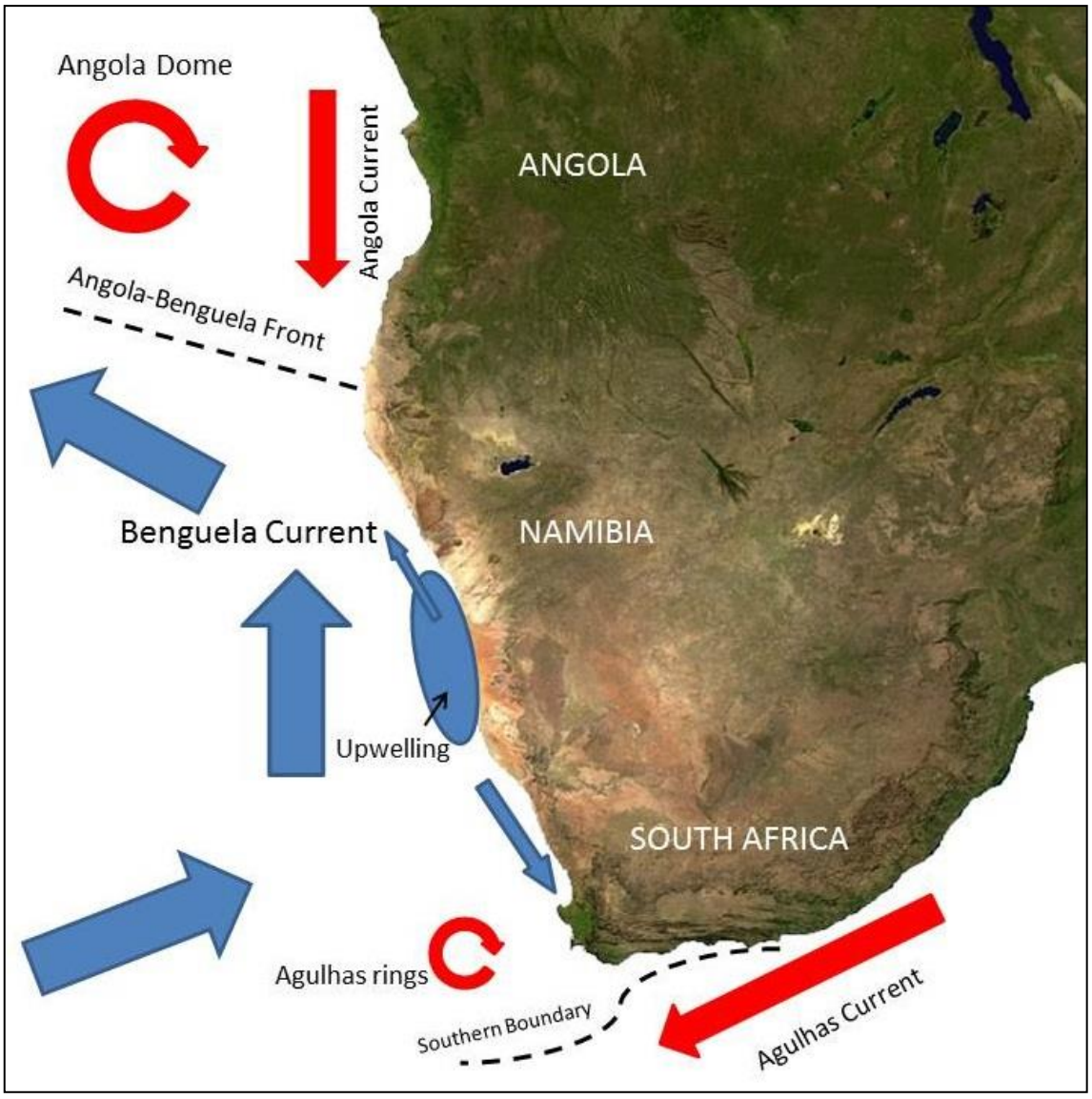

Figure 1.1 The major oceanographic features of the Benguela Current Large Marine Ecosystem, including currents (arrows), oceanographic fronts (dotted lines), and the perennial upwelling cell off southern Namibia (oval). Note the warm Angola and Agulhas currents bounding the Benguela Current to the north and south respectively. Red= warm currents, Blue= cold currents.

The Benguela Current Large Marine Ecosystem (BCLME) is located off the southwest coast of Africa between 15°S and 34°S encompassing the coastlines of southern Angola, Namibia and western South Africa, extending from the Angola-Benguela Frontal Zone in the north to the western Agulhas Bank in the south (Shannon, 1985; Figure 1.1). The Benguela Current is a cold current that forms part of the eastern limb of the subtropical South Atlantic Gyre (see Figure.1.1; Shannon *et al*., 1983). The Benguela Current itself is a highly productive system (Shannon and O'Toole, 2003), with long-shore winds driving numerous upwelling zones from the Angola-Benguela Frontal Zone to the Cape of Good Hope (Hutchings *et al*., 2009). There are three major perennial upwelling cells present in the Benguela Current: Cape Frio (Namibia), Cape

Point (South Africa) and the largest upwelling at Lüderitz in southern Namibia (Hutchings *et al*., 2009; Figure 1.1). The Lüderitz upwelling is the largest and oldest perennial upwelling cell on earth, and divides the Benguela system into northern and southern subsystems (Lutjeharms and Meeuwis, 1987; Shannon and O'Toole, 2003; Demarcq and Dagorne, 2011). The Lüderitz upwelling cell also exhibits cooler average sea surface temperatures and extends further offshore than other upwelling cells in the world (Lutjeharms *et al*., 1991). The scale of the Lüderitz upwelling cell thus makes it a substantial potential barrier to dispersal by tropical and warm-temperate marine taxa (Demarcq and Dagorne, 2011). The Benguela Cold Current system is unique in the world in that it is bounded by two warm currents, the tropical Angolan Current to the north and the warm Agulhas Current to the south: where these cold and warm currents meet forming warm-temperate confluence zones (Shannon and Nelson 1996; Veitch, 2006). Periodically the northern Benguela system is destabilised by the Benguela Niño, a phenomena similar to its Pacific counterpart El Niño. The Benguela Niño, is characterised by a slab of warm, nutrient poor water (up to 50m deep and extending 150km offshore) that enters the northern part of the Benguela upwelling system off the Namibian coast about once every ten years (Florenchie *et al*., 2003) This warm, salty water from the Angola Current moves southward, from 15°S to as far as 25°S. The Benguela Niño has been documented as having detrimental effects of stocks of Sardine (*Sardinops sagax*) in Namibia (Boyer *et al*., 2001), Hjort *et al*., (2012) in a report for the FAO identify Benguela Niño's are likely continue to have a detrimental impact on fisheries in the region.

Palaeoceanographic studies identify the Benguela Current forming around 10 Ma and becoming fully developed around 6Ma (Siesser, 1980; Diester Haass *et al*., 1988; Diester Haass *et al*., 1990; Krammer *et al*., 2006; Jung *et al*., 2014). The uplift of the Isthmus of Panama and the onset of the glacial cycles at the Pliocene-Pleistocene transition (around 2 Ma) led to further intensification of the Lüderitz and other associated upwelling cells, and the establishment of the present day environment of the Benguela Current system (Marlow *et al*., 2000; Krammer *et al*., 2006). The Mid Pleistocene Transition (MPT) from 1.2 Ma- 600 Kya involved the transition from 41 Kyr to 100 Kyr glacial-interglacial cycles during the MPT. Before the MPT the climate was dominated by comparatively warmer glacial periods and cooler interglacials, and sea levels were higher during glacial periods (approximately 70m lower than present

day) than in the subsequent 100Kyr glacials. During the MPT the Benguela Current was more unstable (Marlow *et al*., 2000; Clark *et al*., 2006). However at the end of the MPT (850-750 Kya) there was a strong cooling of the Benguela Current and an overall intensification of currents and upwelling cells, which have remained in place during the subsequent glacial and interglacial periods to the present day (Marlow, 2000). During glacial periods the perennial upwelling cell at Lüderitz has increased intensity (Marlow *et al*., 2000). Further palaeoceanographic studies have identified a larger and increased upwelling activity in the Western Cape region in South Africa (Petrick *et al*., 2015), indicating a stronger and (geographically) more extensive Benguela Current during glacials. The Agulhas Current is greatly reduced in both strength and geographical extent during glacial periods (Peeters *et al*., 2004). During the last glacial maximum sea levels dropped approximately 120-140 m compared to the present day (Ramsay and Cooper, 2002). Accordingly large areas of the continental shelf were exposed above the sea and in South African waters this drop in sea level left the Agulhas bank exposed, extending the southern tip of Africa (Cape Agulhas) 200km further south than its present day locality (Teske *et al*., 2013). During historical interglacials conditions were similar to those of the present day as outlined above and Figure 1.1 with the exception that Agulhas leakage into the Atlantic Ocean was greater than in the present day (Peeters *et al*., 2004).

The Benguela cold upwelling system is thought to be one of the factors promoting faunal differences between the Atlantic Ocean and Indo-Pacific Ocean following the closure of the Tethys Ocean due to the system presenting a physical barrier to dispersal of coastal marine organisms (Floeter *et al*., 2008). However few studies have focused specifically on the Benguela system and its effects upon local fish species. It could be presumed that populations of fish species living either side of the upwelling, being adapted to the warmer temperate to subtropical coastal waters, would find the cold water of the upwelling a barrier to dispersal, thus resulting in a breakdown of gene flow between the isolated populations and the development of genetic differentiation of populations to either side of the upwelling system.

In Spalding *et al.*'s (2007) study of the Marine Ecoregions of the World (MEOW's) all of the southern African marine region was designated as being wholly temperate. However, such a designation ignores the complexities within the region, i.e. the interactions of the warm tropical Angolan Current and sub-tropical Agulhas Current

with the cold Benguela Current and the associated fronts and warm temperate areas between these major current systems. Potts *et al*. (2015) utilised sea surface temperatures (SST's) and distributions of fish species to propose more refined biogeographic regions for coastal fauna around southern Africa. Potts *et al*. (2015) identified multiple regions (see Figure 1.2): a tropical region northwards from the South African-Mozambique border; a subtropical region in north-eastern South Africa (Kwazulu-Natal Province); a warm temperate region along the Eastern and Southern Cape provinces of South Africa; a cool temperate region from False Bay (South Africa) to Henties Baai (Namibia), including the very cold perennial upwelling cell in southern Namibia; a warm temperate region from Henties Baai (Namibia) to Namibe (Angola); a sub-tropical region from Namibe to Rio Longa (Angola); and a tropical region from Rio Longa northwards. The present study therefore aims to use this complex oceanographic region to examine population structuring and gene flow (dispersal) potential of warm temperate coastal fish species whose distributions are bounded by tropical and cold water systems over relatively small spatial scales.

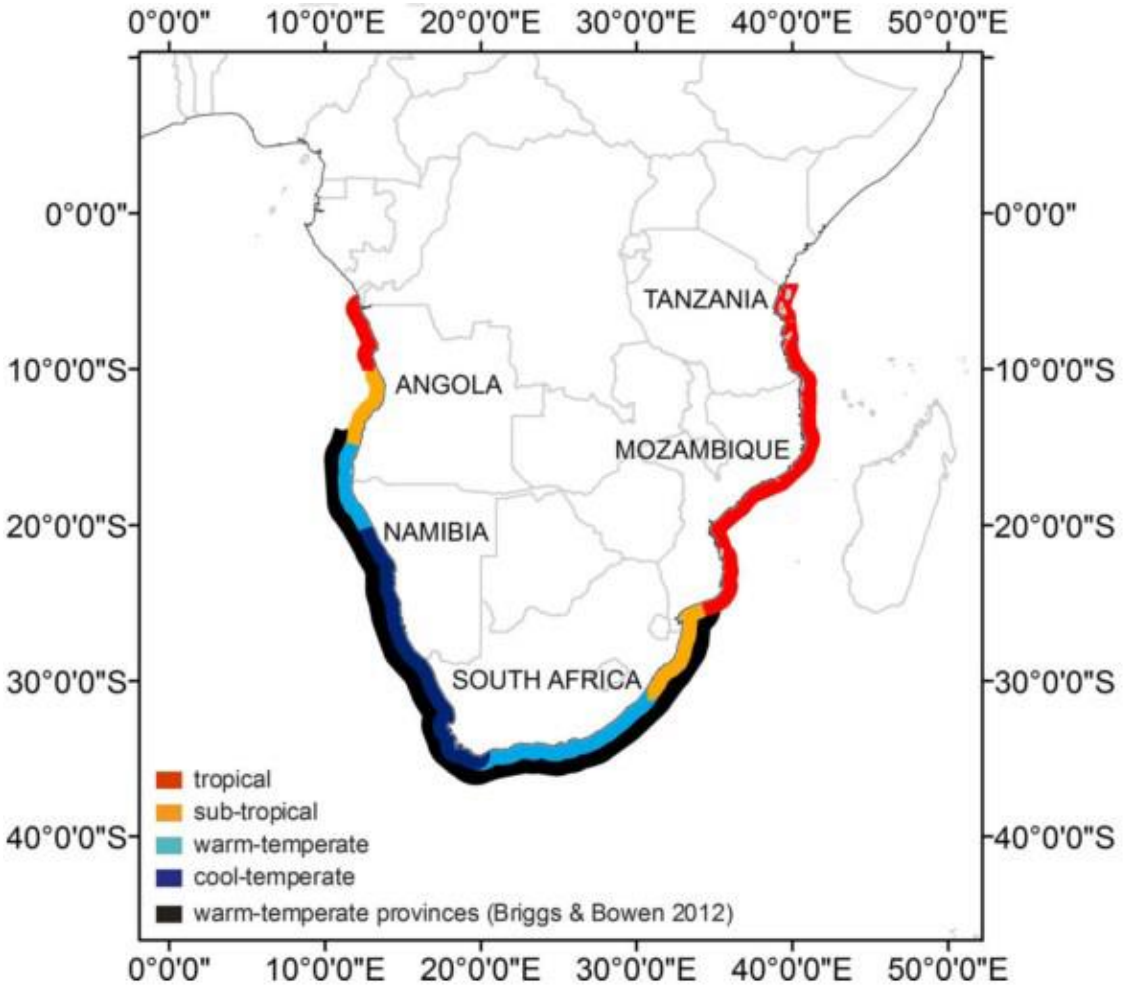

Figure 1.2 Biogeographic zones of the coastal region of southern Africa as presented by Potts *et al.* (2015)

The Benguela Current system may represent a barrier to fish movement and dispersal, and so generate population structuring in several ways. Firstly the cold upwelling at Lüderitz and the cold north-flowing Benguela Current may provide unsuitable habitat for warm temperate and sub-tropical species, resulting in breaks in adult population distribution and unsuitable conditions for larval survival and dispersal (Lessios *et al*., 2003; Lett *et al*., 2007; Hanel and Tsingenopoulos, 2011). Secondly, strong currents could affect dispersal of coastal species by either asymmetric dispersal in the direction of the current (i.e. northwards) or by carrying larvae offshore into unsuitable habitat (the upwelling areas and the offshore deflections from the Agulhas and Angolan Currents). Finally, oceanographic fronts and eddies may promote or inhibit dispersal thereby decreasing or increasing observed population structure (Teske *et al*., 2011).

Whilst the Benguela Current system may act as a barrier to dispersal and a biogeographic boundary it may not be impermeable. Floeter *et al.* (2008) identified 47

contemporary fish species that are thought to have colonised from the Indian Ocean into the Atlantic Ocean, 38 of which are found in the Tropical East Atlantic (TEA; i.e. the west coast of Africa north of the Angola-Benguela front and south of Mauritania). Colonisation into the TEA from the Indian Ocean is believed to occur via sporadic warm cyclonic eddies ('Agulhas Rings') which spin off the end of the Agulhas Current off south-western South Africa and traverse the Benguela Current (Penven *et al*., 2001; Peeters *et al*., 2004; Briggs and Bowen, 2013). The system may thus promote sporadic dispersal through the Benguela system from south to north.

Most research on the Benguela system has focussed upon the southern subsystem in South Africa. This is largely due to previous inaccessibility of Angola due to the civil war which ended in 2002, and large sections of the coastline in western South Africa and Namibia being closed to researchers due to diamond mining. Olivar and Shelton (1993) studied larval fish assemblages across the Benguela region, finding larval assemblages were least diverse (in species) in the central cold upwelling area. Olivar and Shelton (1993) also found that several species distributed across the region spawned only in the warmer waters associated with the Angolan and Agulhas Currents, suggesting that the upwelling and its associated cold waters present a barrier to dispersal in larvae of coastal fish species.

Evidence showing population structure in marine species (largely warm temperate) that live in the southern Benguela subsystem is limited. West-coast rock lobster (*Jasus lalandii)* exhibited no genetic population structuring across the southern subsystem (Matthee *et al*., 2007), and neither did the goby (*Caffrogobius caffer;* Neethling *et al.*, 2008) or the Bluntnose Klipfish (*Clinus cottoides;* von der Heyden *et al*., 2008). Likewise, Matthee *et al.* (2006) found that cape fur seals (*Arctocephalus pusillus pusillus*) exhibited no noticeable population structure, but suggested that population sizes were likely to have been largest during interglacial periods when the upwelling systems are postulated to be most productive. Von der Heyden *et al*. (2007 and 2010) found that cape hake (*Merluccius capensis*) exhibited a panmictic population whilst *Merluccius paradoxus* exhibited significant population structuring across the southern subsystem. Von der Heyden *et al*. (2010) also described high mtDNA haplotype diversity but low nucleotide diversity in *M. paradoxus*, a pattern typical of a population expansion after a bottleneck (where there are initially few haplotypes in the population). This pattern of expansion has been found to be typical for South African marine animals

where many species show recent expansions starting from ~18 Kya, at the end of the last glacial maximum (Gopal *et al*., 2006; Matthee *et al*., 2006; Von Der Heyden *et al*., 2007 and 2010; Henriques, 2012).

To fully evaluate the role that the Benguela Current system has had and is having on coastal fish species, studies which explore the system in its entirety are required, i.e. including groups of species distributed across both southern and northern subsystems. A handful of studies have been conducted so far, and already describe a highly variable picture among species from evidence of speciation-level effects to very limited within-species population structuring. Henriques *et al.* (2012) confirmed allopatric genetic divergence of two recognised Kob species between the two subsystems, the Dusky Kob (*Argyrosomus japonicas*) in South Africa and the West Coast Dusky Kob (*A. coronus*) in southern Angola, identifying the Benguela Current to be acting as both a present day and historical dispersal barrier leading to speciation from an initial isolation dating to around 2 Ma. Henriques *et al*. (2014) also identified two genetically highly divergent Geelbeck (*Atractoscion aequidens*) populations associated with the two subsystems. In a subsequent study Henriques *et al*. (2016), utilising a combination of mtDNA, nuclear genetic markers, morphology and life history traits, confirmed the genetic divergence in *A. aequidens* is likely a cryptic speciation event dating to around 2-3 Ma. For both speciation events divergences were dated to around 2 Ma, during the Pliocene-Pleistocene transition, creating divergent populations in the northern (i.e. Angola) and southern (i.e. South Africa) warm temperate regions, a timing that coincides with the proposed onset of the present conditions in the Benguela system around 2 Ma.

Whilst the Benguela Current can act as a barrier in some species it can also present a novel habitat for other warm temperate fishes to colonise and adapt to. Henriques *et al.* (2014) identified a genetically divergent population of *Argyrosomus inodorus*, generally considered a warm temperate species found throughout South African coastal waters, living within the Benguela Current in northern Namibia. In an associated study Potts *et al.* (2014) identified potential hybridisation between *A. inodorus* and the sister species *A. coronus*, likely as a direct result of rapid warming within the region leading to secondary contact between the two species.

Whilst some coastal taxa show evidence of allopatric speciation, some taxa studied across the Benguela Current system display significant genetic divergence at the level of intra-species structuring. Henriques *et al*. (2012) identified genetically divergent

populations of Leervis (*Lichia amia*) across the Benguela Current with a date of divergence of around 200 Kya. Likewise Henriques (2012) identified significant genetic structuring across the Benguela Current in the Blacktail Seabream (*Diplodus capensis*) with divergence dating to around 360 Kya. Such variable levels of divergence suggest that whilst the Benguela Current has been a potentially impermeable barrier to dispersal since the establishment of its present oceanographic structuring 2 Ma, for some coastal fish it has been permeable periodically during this period.

The Benguela Current system also has been found to be a prominent barrier to dispersal for other non-teleost taxa. The shark *Triakis megalopterus* displayed genetic differentiation across the Benguela Current associated with differing life history traits: Angolan *T. megalopterus* grow faster, breed at an earlier age and have a shorter lifespan than their South African counterparts (Soekoe, 2016). de Beer (2014) found populations of *Octopus vulgaris* to exhibit significant genetic divergence yet no morphological variation.

1.5 Overall project structure and aims

The present project aims to investigate the interaction between oceanographic features, palaeoceanography, life history features and genetic structuring and evolution of marine coastal fishes. The study will focus on genetic diversity of a range of Sparid fish species distributed across the Benguela Current region, but also investigate the evolutionary relationships of these species to con-specific populations or closely related sister species distributed across the west coast of Africa, the northeast Atlantic and the Mediterranean. Finally, the project will aim to synthesise all available genetic data of the fish species studied during this and previous projects (principally Henriques, 2012) to establish any emerging patterns or commonality in how coastal fish populations are structured and evolving in the Benguela Current Large Marine Ecosystem (BCLME).

To investigate which features of the Benguela Current system may act as barriers to dispersal and genetic connectivity, or which species' life history features combine with environmental variables to result in different levels of isolation and breakdown in gene flow, the study attempted to cover species (or species complexes) with distributions in both the northern and southern Benguela subsystems. The study species chosen were:

1. *Spondyliosoma cantharus* and *Spondyliosoma emarginatum.* This species pair is distributed to either side of, and apparently isolated by, the Benguela system. *S. cantharus* distributed from northern Namibia to the NE Atlantic and Mediterranean and *S. emarginatum* distributed from western South Africa to north-eastern South Africa. For an introduction to species biology and distributions see Chapter 2.

2. *Sarpa salpa* and *Lithognathus mormyrus*. These two species show similar distributions that span from the NE Atlantic / Mediterranean to north-eastern South Africa, with a distribution break associated with the cold water regions of the Benguela Current system. For an introduction to species biology and distributions see Chapter 3.

3. *Diplodus cervinus/ hottentotus*. This species complex has unresolved taxonomy, and there are proposed boundaries to populations / subspecies associated with oceanographic features found within the Benguela Current system and between the Mediterranean and NE Atlantic. For an introduction to species biology and distributions see Chapter 4.

1.5.1 Key issues and questions to be resolved

I. Establish patterns of phylogeography and genetic population structuring for the study species across their ranges from southern Africa and Europe.
II. Investigate the role of past climate or geological events in shaping present day population structuring.
III. Establish which present day oceanographic features are most important in shaping genetic connectivity and population structuring.
IV. Use genetic methods to help resolve species validity and taxonomy.

A wider application for the project is that all of the study species are of commercial or artisanal importance to local fishing communities throughout their ranges. Thus data and results generated in this project will be used, in conjunction with associated development projects active in the region, to help sustainably manage the marine resources of the BCLME. The fisheries in Angola remain underdeveloped, particularly when compared to South Africa, whilst artisanal fisheries in the region remain an important food source, particularly so for Angola (FAO, 2014), although some fish species are considered to be overfished in Angola (FAO, 2014). As such Angola

presents an opportunity to develop a sustainable well managed fishery. Whilst South African fisheries are developed with management systems in place they are continually further developed with management and conservation plans being reviewed (Hauck and Croese, 2006). The present study will support fisheries management and conservation in the region by aiding species identification, evolutionary history, definition of stocks and / or assessment of population structure. For example rather broadly this study will identify whether species with trans-boundary distributions (i.e. occur in both Angola and South Africa) represent distinct stocks. Where relevant this study will outline fisheries and conservation management considerations.

Chapter 2: Genetic evidence for cryptic speciation across marine biogeographic boundaries in the Sparid genus *Spondyliosoma*.

2.1 Introduction

While many marine species conform to traditional models of high gene flow and weak population structure over large spatial scales (Palumbi, 1994; O'Reilly *et al*., 2004), there are a growing number of studies revealing structuring on local geographic scales (Hauser and Carvalho, 2008). In many cases genetic divergence is often driven by localised areas of reduced gene flow, i.e. barriers (Cowen *et al*., 2000). Common barriers include transient physical barriers due to sea-level changes and oceanographic features (eddies, currents and upwelling zones; Cowen *et al*., 2000), as well as regions of ecological clines that may promote the coupling of endogenous and exogenous gene flow restrictions.

Within the East Atlantic two intense upwelling zones have been described: off the coast of Morocco / Mauritania; and the Benguela cold upwelling(s) on the border of South Africa and Namibia. Phylogeographic studies have revealed strong signatures of genetic divergence in marine taxa coinciding with these upwellings (e.g. studies of East Atlantic: Chikhi *et al*., 1998; Durand *et al*., 2005). Studies of southern African coastal fishes have in many cases reported divergence between samples from South Africa and those from north of the Benguela upwelling (Teske *et al*., 2011; Henriques *et al*., 2012, 2014, 2015). For example, the Benguela Current region has been implicated in the evolutionary diversification of two sister species of Croaker, *Argyrosomus japonicus* in South Africa and *A. coronus* in Angola (Henriques, 2012). The level of divergence and timing of the split between these *Argyrosomus* sister species is estimated to be concurrent with the intensification and onset of the present day oceanographic features of the Benguela Current during the Miocene-Pleistocene transition around 2Ma. Whilst the Benguela Current appears to act as an impermeable barrier to movement and dispersal for some coastal fishes, for others it presents a semi-permeable barrier which

at least on historical timescales can be traversed and some gene flow effected (e.g. *Diplodus capensis* - Henriques *et al*., 2012)

2.1.1 Study species - *Spondyliosoma spp.*

Within the Sparid genus *Spondyliosoma* the Benguela Current represents a biogeographical boundary for the two described species, *Spondyliosoma cantharus* (Black Seabream) and *S. emarginatum* (Steentjie). *S. cantharus* is described in coastal waters throughout the North East Atlantic and Mediterranean Sea; whilst *S. emarginatum* is regarded as endemic to southern Africa (Figure 2.1). The southern limit of *S. cantharus* is unknown but presumed to be somewhere along the west African coast, probably between Angola and South Africa with the Benguela Current potentially acting as a boundary to further dispersal (Figure 2.1)**.**

S. cantharus and *S. emarginatum* are demersal spawners, and both species exhibit interesting and relatively complex breeding behaviours (e.g. nesting and nest guarding) compared to many Sparids and coastal fishes in general. Nesting sites tend to be located on thin gravel beds covering bedrock, in shallow waters approximately 5-10m deep. Breeding males have a humped shoulder, concaved forehead and during the breeding season are dark silver to black in colour with an iridescent blue- grey band between the eyes. Males use their tails to remove the loose surface layer of the gravel bed exposing bedrock or more compacted gravel thus excavating a nest. The ideal nest size is between 1-2m^2(Southern Science, 1995). Females then lay their sticky eggs in a thin layer within the nest where the eggs become strongly attached to the rock/gravel surface. It is suggested that males use their nest in male-male competition to attract females (Pawson, 1995). The male then guards the nest for 2-3 weeks until the eggs hatch. Juvenile *S. cantharus* remain in the nest environment until they are 7-8 cm in length. After this the juveniles disperse locally, remaining in the inshore area for 2- 3 years (until approximately 20cm in length) when they become sexually active and join the adult population (Pawson, 1995). Demersal spawning in nests with juveniles remaining close to natal nests during early development would be expected to severely limit dispersal by early life stages in these species compared to other coastal fishes with a pelagic egg or larval phase.

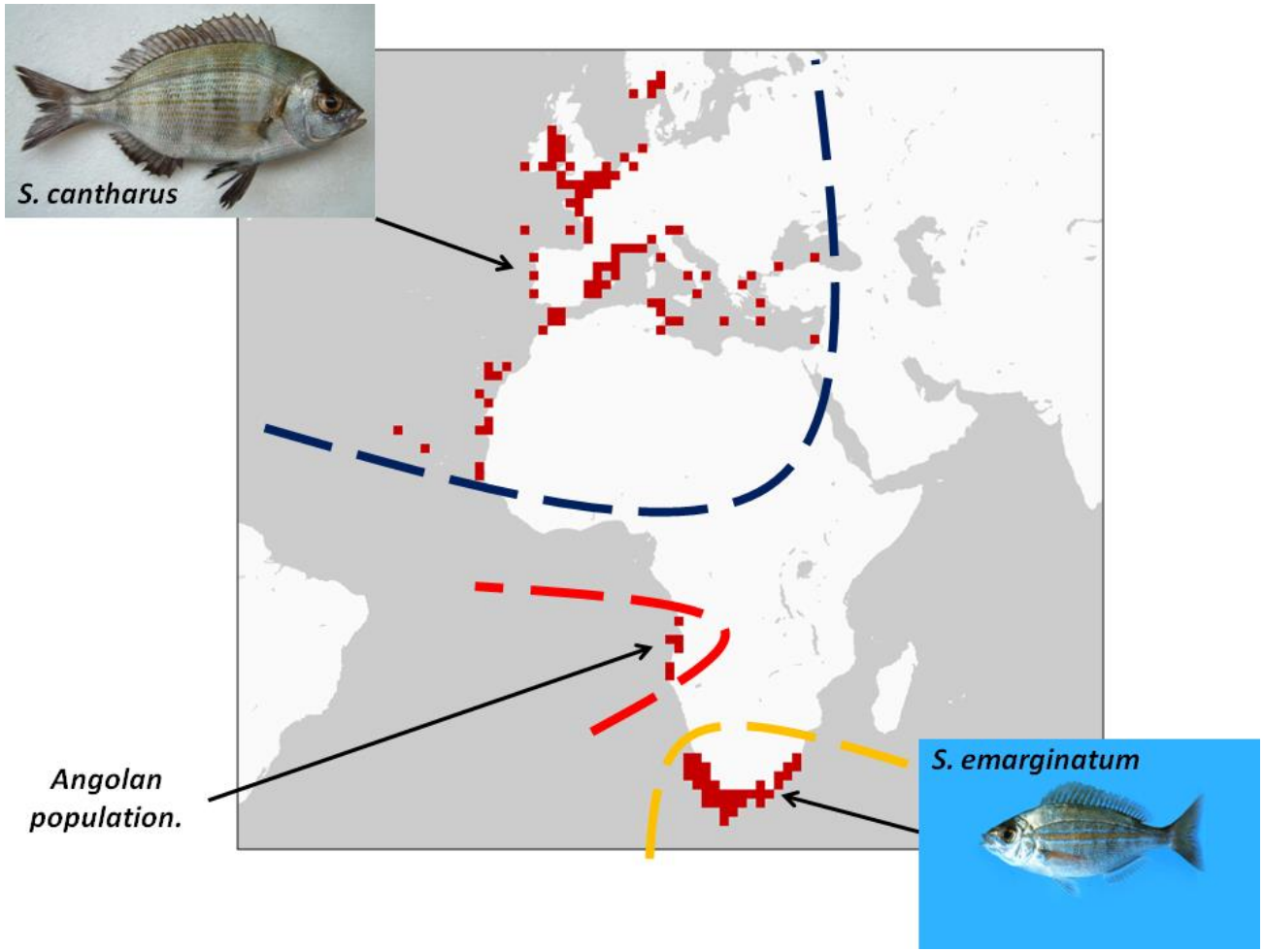

Figure 2.1 Geographical distribution map for the two described *Spondyliosoma* species and the isolated Angolan population of unknown identity. Distribution map from GBIF, available at: http://www.gbif.org

Contrary to the early life stages of *S. cantharus*, adults exhibit considerable mobility and dispersal potential, which would be expected to promote gene flow. In the English Channel *S. cantharus* are known to exhibit migratory behaviour, overwintering in deep water (50-100m) west of a line from Start Point (England) to Alderney (Channel Islands) then in spring appearing to follow the 9°C isotherm moving northeast into the eastern English Channel (Pawson, 1995). Breeding occurs off the southeast coast of England from May- July, with adults then moving further north into the North Sea before retreating back down the Channel to overwinter again in the deep water of the Western Approaches (Pawson, 1995). In Cardigan Bay (Irish Sea) *S. cantharus* appear to make a similar migration travelling northwards during spring (presumably following the 9°C isotherm) before breeding from late May into July. *S. cantharus* in Cardigan Bay are known to return to the same nesting sites year upon year, and local fishermen have noted a trend of nesting sites occurring further north and increasing in abundance since the 1970s (pers. comm.). Return to nesting sites year after year could indicate a certain degree of philopatry, which could act to increase population structuring by reducing dispersal. Such migratory behaviour has not been studied (or observed) in the rest of the range of *S. cantharus* or in *S. emarginatum*. *S. cantharus* is near its most northern range limit in UK waters, so it is possible that such migrations occur as the winter temperatures in Cardigan Bay and the eastern English Channel are below the thermal tolerance of adults, whilst further south winter temperatures are high enough to allow adults to be resident all year and so reducing potential dispersal further. Alternatively *S. cantharus* may over-winter in the Western Channel due to an abundance of prey items, as cuttlefish *(Sepia officinalis)* juveniles are known to overwinter in the western channel and are actively preyed upon by *S. cantharus*. In summary adult *S. cantharus* (and possibly *S. emarginatum*) display considerable dispersal potential which may promote panmixia, but which would be restricted if adults show philopatric migrations throughout the range, whilst the larval stages are demersal and appear to display limited dispersal capability.

There have been no formal taxonomic (morphological, life history or genetic) comparisons between the two species, with the exception of the larval stages. Juvenile *S. cantharus* exhibit melanophores on the head at 3.2 mm length and on the caudal fin and lower jaw by 5.3mm length, all of which are absent in *S. emarginatum* juveniles; *S. cantharus* also complete the flexion larval stage earlier than *S. emarginatum* (Ranzi,

1933; Russell, 1976; Beckley, 1989). *S. cantharus* has an elliptical compressed body shape with a single dorsal fin which is spiny rayed at the front, a forked tail and a relatively small mouth (Miller and Loates 1997; Lythgoe and Lythgoe 1971; See Figure 2.1). Adults are coloured silver with a bluish tinge. Juveniles often have broken latitudinal golden lines, which can persist in some individuals into adulthood (Miller and Loates 1997; Figure 2.1). The common length for *S. cantharus* is 30 cm SL, with a maximum reported body size of 60 cm SL Male (Bauchot and Hureau 1986). *S. emarginatum* is morphologically similar to *S. cantharus*, but appears to be smaller than *S. cantharus* with the maximum reported size being 45cm TL and a common adult size of 25cm TL (Figure 2.1; Bauchot and Smith 1984; Smith and Smith 1986).

Most recent research suggests that both *S. cantharus* and *S. emarginatum* are protogynous hermaphrodites, first sexually active as females and then maturing into males (Pajuelo and Lorenzo, 1999; Goncalves and Erzini 2000; Mouine *et al*., 2010). Sexual maturity occurs at around 20cm TL (approximately 2 years old), with sex change from female to male occurring at around 30 to 40 cm TL (around 3 years old) in *S. cantharus*. Therefore larger individuals within a population tend to be males. However there is considerable variability regarding the size at sex change in *S cantharus*, with Pajuelo and Lorenzo (1999) reporting *S. cantharus* in the Canary Islands mature as females at 17.3 cm TL (2 years old), and change sex at 22.7cm TL (3 years old): this reduced size is possibly indicative of overfishing of *S. cantharus* in the Canary Islands. The hermaphroditic nature of *S. cantharus* and *S. emarginatum* may have consequences for their susceptibility to overfishing fishing efforts in UK waters often focus on the larger *S. cantharus* individuals (i.e. breeding / nest guarding males), during their breeding season. Such a focus may have an effect of biasing the sex ratio of populations to females thereby affecting reproduction and repopulation of stocks. For example between 1977 and 1978 the modal size of the *S. cantharus* UK catch decreased from 37-38cm to 28-30cm as the fishery expanded (Pawson, 1995).

S. cantharus was included in a study to identify Mediterranean Sparids by Isozyme markers (Alarcon and Alvarez, 1999). Presently there has only been one study into the population genetics and phylogeography in *Spondyliosoma* using DNA markers. Bargelloni *et al*. (2003) studied the Mediterranean and North East Atlantic proportion of the *S. cantharus* range, utilising a combination of nuclear allozyme loci and 180bp mtDNA Control Region (CR) sequences. The CR data in the Bargelloni *et al*. (2003)

study identified a pronounced genetic break between Mediterranean and Atlantic *S. cantharus*, but identified no genetic structuring between sites within the Mediterranean. A re-evaluation of the Bargelloni *et al.* (2003) data set by Patarnello *et al.* (2007) reaffirmed the previous finding and also identified that the *S. cantharus* Mediterranean population has been historically stable. So whilst the North East Atlantic and Mediterranean range of the genus has been previously studied there remains a need for a phylogeographic study of the remainder of the range occupied by the genus (i.e. Angola and South Africa). The IUCN Red list of endangered species list both species as being of least concern however it is recommended that efforts should be undertaken to protect both species from overfishing (Mann *et al*., 2014; Russell *et al.*, 2014c). Such a study should thus help inform future conservation and stock management efforts.

2.1.2 Aims and objectives.

This study employs multiple genetic markers (mtDNA Cytochrome Oxidase I (COI) and CR sequences, plus nuclear microsatellites) to apply a combination of phylogeographic and population genetic analyses to address the following questions:

1. Is there genetic evidence for the two recognised species *S. cantharus* and *S. emarginatum*? If so we would expect both species to form reciprocally monophyletic clades in reconstructed phylogenies, corresponding to the recognised species and distributions.
2. Are there any further major phylogenetic or phylogeographic subdivisions within *Spondyliosoma*? Bargelloni *et al.* (2003) identified a pronounced phylogeographic break between Atlantic and Mediterranean *S. cantharus* utilising CR: is this break still identifiable with more conservative genetic markers such as COI?
3. What is the demographic-genetic status of *Spondyliosoma* found in Angolan waters? The Angolan population appears to be isolated from the rest of the species/genus range (Figure 2.1), but this population has been assumed to be *S. cantharus* with the Benguela Current representing the major barrier separating it from *S. emarginatum* in South Africa. The species identity of this population thus remains unclear as to whether it represents the most southerly population of *S. cantharus* or the most northerly population of *S. emarginatum* (or something else?).

4. Is there evidence for additional genetic sub-structuring that may implicate a finer scale population structuring within regions? Given the contradictory picture of low early stage dispersal and potentially high (migratory) adult dispersal, it could be predicted that either distinct geographical structuring or region-wide panmixia (respectively) might result.

5. Do the sampled populations exhibit a pattern of demographic expansion indicating past population size changes? Many coastal fish species in the Benguela Current region exhibit population contractions / expansions corresponding with past glacial cycles: do *S. cantharus* and *S. emarginatum* display signals of similar demographic changes?

2.2 Methods

2.2.1 Sampling and DNA extraction

Samples of *Spondyliosoma* were collected from eight Angolan and South African locations (Table 2.1). Additionally *S. cantharus* samples were obtained from Cardigan Bay (Wales) to act as confirmed specimens of *S. cantharus* and European Atlantic outgroup for analyses. Samples were obtained from a mixture of recreational angling and local fish markets. A fin clip was removed from each individual and preserved in 95% ethanol. Total genomic DNA was extracted following the phenol-chloroform method described by Sambrook *et al*. (1989) and visualised on a 1% agarose gel.

Table 2.1 Sampling strategy for *Spondyliosoma* spp. with numbers of individuals successfully genotyped for mtDNA COI, CR, and four nuclear microsatellites. * Denotes COI sequences derived from GenBank.

Region	Locality (ID)	Sample size	COI	CR	Microsatellites
Angola	Benguela (BEN)	29	13	9	27
	Lucira (LUC)	15	8	5	-
	Namibe (NAM)	30	11	12	29
	Tombua (TOM)	15	9	5	-
South Africa	Langebaan (LAN)	-	1*	-	-
	False Bay (FAL)	30	9	13	27
	Mossel Bay (MOB)	21	9	15	21
	Tsitsikamma (TSI)	9	1*	6	-
	Port Elizabeth (PEL)	6	-	6	-
	KwaZulu-Natal (KZN)	-	5*	-	-
North East Atlantic	Aberystwyth (ABE)	16	15	16	-
	Portugal (POR)	-	16*	-	-
Mediterranean	Turkey (TUR)	-	21*	-	-

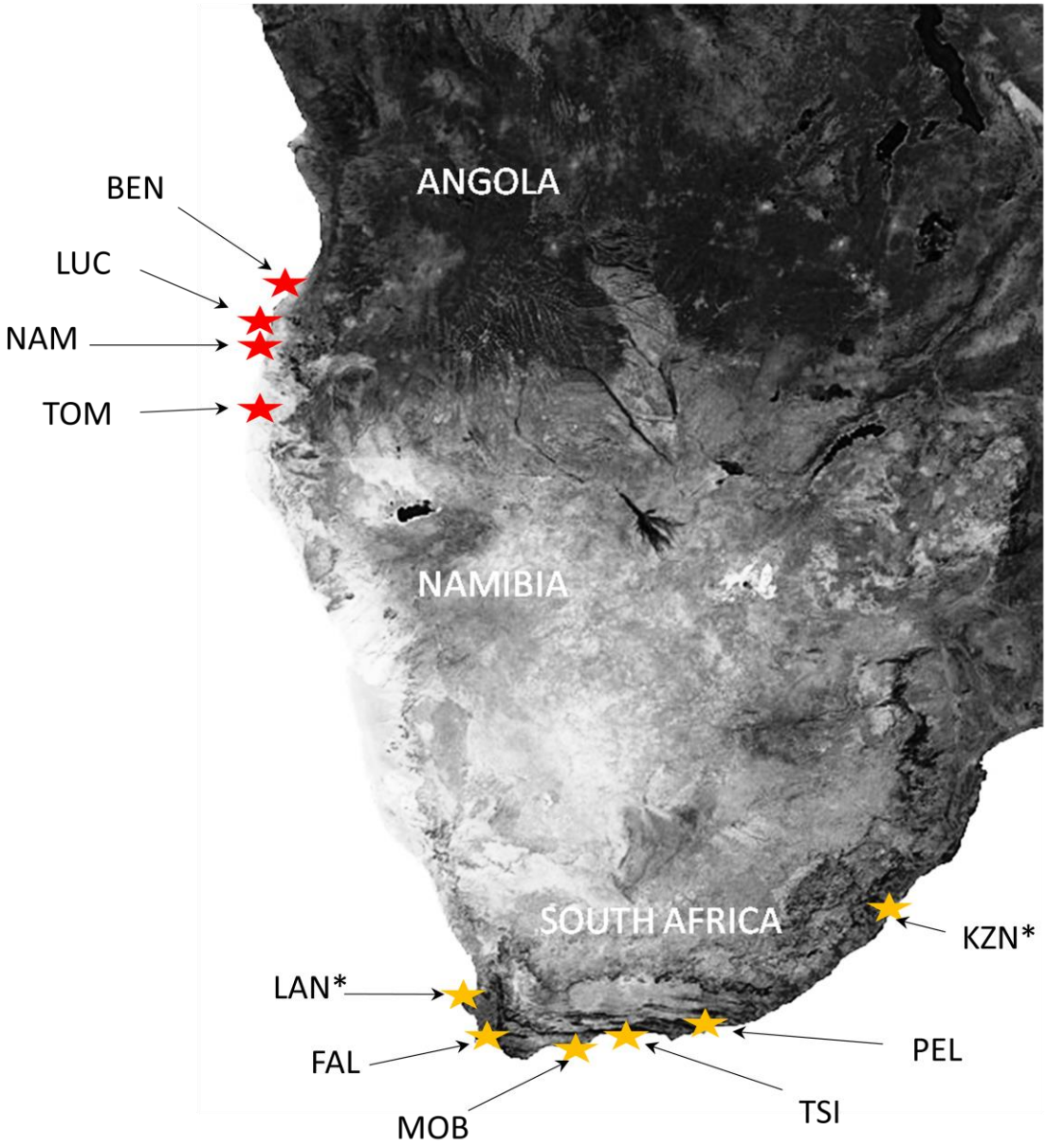

Figure 2.2 Southern African sampling strategy for *Spondyliosoma*. Sampling Site ID codes are as per Table 2.1. * denotes sampling site comprising COI sequences derived entirely from GenBank.

2.2.2 Mitochondrial DNA markers

Sequence variation was assessed in two mitochondrial DNA (mtDNA) regions: COI CR. The COI region was amplified by PCR using the universal fish primers COI-WF1 and COI-WR1 (Ward *et al*., 2005). Owing to a high number of PCR failures using the universal primers, the generated sequences (and GenBank sequences) were used to design new primers using PRIMER3 0.4.0 (Rozen and Skaletsky, 2000), generating the following primers: SCCOIF: 5' GCTTGAGCCGGAATAGTAG 3' and SCCOIR: 5' TTGGTAAAGAATTGGGTCTCC 3'.CR was initially amplified by using the universal CR primers L-PROF (Meyer *et al.,* 1994) and 12SAR-H (Palumbi, 1996), with preliminary sequences used to develop species-specific primers: Forward: 5' CACACATAATGTTAGAGATATAGGA3' and Reverse: 5' TGCATAAGTGATTTCATGAGCATAAT 3' which were predicted to amplify a 436bp fragment of the first hyper variable region of the CR.

PCRs comprised of 10 µl of BIOMIX (BioLine), 1.0 pMol of primer (both forward and reverse), 6 µl of template DNA and 2 µl of sterile distilled water giving a total reaction volume of 20µl. All PCRs were performed using a C1000 Thermal Cycler (Bio-Rad) using the following reaction conditions: COI – 120 s at 95°C, then 40 cycles of 30 s at 94°C, 30 s at 50°C, 60 s at 72°C, with a final extension step of 120 s at 72°C; CR – 300 s at 94°C, then 30 cycles of 30 s at 94°C, 60 s at 51°C, 120 s at 72°C. These reaction conditions were identified as being optimal for both species. Resultant products were then verified using a 2% agarose gel stained with Gel Red (PROMEGA). PCR products were purified by use of SureClean (BioLine). PCR products were sequenced using an ABI 3730 DNA analyser (Applied Biosystems®).

2.2.3 mtDNA data analysis

Sequence chromatograms were edited manually using Chromas Lite (Technylesium Pty Ltd) and aligned using BioEdit (Hall, 1999). Sequences were prepared for analyses using DAMBE (Xia, 2001) and DnaSP v5 (Librado and Rozas, 2009).

Genetic variation was quantified by standard indices of haplotype number (H), haplotype diversity (h) and nucleotide diversity (π) calculated in ARLEQUIN 3.5 (Excoffier and Lischer, 2010). These values were calculated for each individual sampling site, sampling region (sites pooled e.g. South Africa) and if identified by subsequent analysis individual clades.

Phylogenetic relationships among haplotypes were inferred using network and tree based methods. Haplotype networks were constructed using a median-joining algorithm in NETWORK 4.6 (Bandelt *et. al.,* 1999). Network based phylogenies have several attributes which make them useful for reconstructing intraspecific and recently diverged taxa's evolutionary relationships, namely they account for multifurcations, presence of both descendant and ancestral haplotypes and recombination events producing reticulate relationships, which are not accounted for by traditional bifurcating trees (Posada and Crandall, 2001; Ferreri *et al*., 2011).

Models of evolution of nucleotide substitution are indispensable when using DNA sequences to estimate phylogenetic relationships among taxa (Posada, 2009). Model selection may affect estimates of branch lengths, branch support values, transition / transversion ratio, underestimate divergence and substitution rates (Posada and Buckley, 2004). Whilst phylogenetic trees inferred without the correct substitution model may be less accurate (recover an incorrect tree more often) or may be inconsistent (converge on an incorrect tree with increased amounts of data; Posada and Buckley, 2004; Posada, 2009). No model can completely relay the true complexity of molecular evolution, despite their simplified assumptions a model can fit the data and make accurate predictions. The best fitting substitution model of the observed sequence polymorphism in this study was estimated using jModelTest 2.1.4 (Guindon and Gascuel, 2003; Darriba *et. al.,* 2012). Computation of likelihood scores for the different models was performed for eleven substitution schemes (88 Models), this included estimating the proportion of invariable sites and the gamma distribution optimised using a Maximum likelihood base tree. Both Nearest Neighbour Interchanges (NNI) and Subtree Pruning and Regrafting (SPR) tree rearrangement algorithms were tested. The most probable substitution model was chosen based upon the Bayesian Information Criterion (BIC) and used in the subsequent analyses.

Maximum likelihood tree construction was performed in PhyML (Guindon and Gascuel, 2003). PhyML uses an algorithm to produce phylogenetic trees by maximum likelihood (ML). The Program creates an initial tree using a distance-based method that it modifies to improve likelihood each time. Trees were improved using two heuristic methodologies: NNI and SPR along with optimising branch length and topology. Branch support values were calculated using bootstrapping with 1000 replicates. Phylogenetic trees are displayed using FigTree (Rambaut, 2014). Outgroup taxa were identified using

the most closely related taxa for which sequences are available. As such *Sarpa salpa*, *Boopsoidea inornata* and *Spicara maena* are used as outgroups for the COI phylogenies, and *S. salpa* and *S. maena* as outgroup taxa for the CR phylogenies.

Net genetic distance between populations was calculated using the minimum p-distance (the proportion (p) of nucleotide sites at which two sequences being compared are different) in MEGA 6 which also provides a standard error value for the estimated distance (Tamura *et al*., 2013). Minimum p- distance is used in the present study as use of mean p-distance can artificially inflate the distance between two taxa, which when considering species delimitation may lead to misclassification (Meier *et al*., 2008). Time since divergence (Da) between significantly differentiated populations was estimated using the formula $t = d/\mu$, where t = time in years, d = net genetic distance between populations and μ = % sequence divergence per million years (Takahata and Nei, 1985). Values used for divergence rate are those typically found in marine teleosts and varied depending on the marker: 1.2% per million years for COI (Bermingham *et. al.* 1997; Bowen *et. al.* 2001; Lessios, 2008; Henriques *et al*., 2014) and 3.6% per million years for CR (Henriques *et al.*, 2014).

To test for population structure pairwise Φ_{ST} tests were performed between each individual sampling site and between regions in ARLEQUIN 3.5. Φ_{ST} was calculated as it incorporates genetic distance rather than relying on haplotype frequency alone (Holsinger and Weir, 2009). Analysis of molecular variance (AMOVA) was performed in ARLEQUIN 3.5 to test the effect of the Benguela Current on genetic structuring of *Spondyliosoma* in southern Africa (i.e. Angolan vs. South African samples).

2.2.4 Demographic analyses

Neutrality tests are known to be sensitive to past population size changes and subsequent deviations from mutation-drift equilibrium. Here Tajima's *D* (Tajima, 1989) and Fu's F_S test (Fu, 1997) were performed in ARLEQUIN 3.5. Both tests are known to be particularly sensitive to historical demographic change (Librado and Rozas, 2009).

History of the effective populations size was investigated using mismatch distribution within regions (North East Atlantic, Mediterranean, Angola and South Africa) analysing the distribution of pairwise differences amongst all haplotypes occurring within regions using models by Rogers and Harpending (1992) and Rogers (1995). Mismatch distribution analysis is based upon the observation that populations undergoing a

population expansion exhibit a unimodal distribution of pairwise differences, whereas populations with stable demography exhibit a multimodal distribution of pairwise differences. Models simulated an expected scenario where the population has undergone a sudden expansion; this model was fitted against the observed pairwise differences. In addition the raggedness index (goodness of fit) was calculated for regions. Dates of past population expansion were estimated with the formula $T = \tau/2u$ (Rogers and Harpending, 1992), where T = time since expansion, τ = a moment estimator, which represents a unit of mutational time and *u* is the cumulative (across the sequence) probability of substitution.

2.2.5 Microsatellite DNA markers

Initially 18 microsatellite loci developed for several Sparid species (Table 2.2) were tested for cross species amplification for *Spondyliosoma* (as well as the other three Sparid taxa studied in this thesis- *Diplodus cervinus*, *Diplodus hottentotus*, *Lithognathus mormyrus* and *Sarpa salpa*). To test for cross species amplification and optimise PCR conditions eight individuals from South Africa (Mossel Bay) and eight individuals from Angola (four Benguela and four Namibe) were used in several PCR's using a gradient Ta from 45-60 °C with the following PCR conditions: 300s at 95°C, then 30 cycles of 30s at 92°C, 30s at gradient Ta, 30s at 72°C and a final extension step of 72°C for 120s. Success of amplification was assessed visually using a 2% Agarose Gel stained with GelRed. If successful amplification was identified subsequent PCRs were performed to verify optimal Ta. Additionally PCR cycle number, amount of primer used, duration of Ta and MgCl2 concentration were varied but ultimately the standard conditions outlined above and below prevailed to be the most optimal regardless of species or loci used with Ta being the most important factor in ensuring amplification under PCR. All reactions used 5 µl of BIOMIX (BioLine), 0.5 pMol of primer (both forward and reverse), 3 µl of template DNA and 1 µl of sterile distilled water giving a total reaction volume of 10µl. Final confirmation of successful amplification was assessed by alleles separation using an AB3730 DNA analyser and allele identity and number inferred using Peak Scanner 2. Following testing and optimisation of the DNA primers a subset of eleven loci which provided consistent PCR amplification success were sent for analysis using the initial 16 individuals. Subsequently only four loci were identified with suitable levels of polymorphism for

subsequent population screening, whilst the remaining amplifiable loci were either monomorphic or failed in the sequencing reaction (and did so upon re- analysis).

The present study was designed to be highly focussed on the area bounding the Benguela Current, with the sampling regime reflecting this (Table 2.1), with multiple samples sites in Angola and South Africa to investigate whether there is any genetic sub-structuring within these regions. As such the microsatellite component of this study (and for species in subsequent chapters) focuses on the finer scale population structuring in the areas bounding the Benguela Current (Angola and South Africa), with the out group individuals (i.e. North East Atlantic) are not included in the microsatellite component since they are not relevant to the finer scale population structuring. Furthermore inclusion of these outgroups in the microsatellite analysis was not feasible given both the financial and time constraints of this study. Samples were amplified using PCR using the above PCR conditions and PCR mix with the optimal Ta used as identified in Table 2.2 for each loci.

Table 2.2 The eighteen microsatellite loci tested in the present study for all four study species, showing: microsatellite ID; the original species the primers were developed for (DS = *Diplodus sargus*, DV = *D. vulgaris*, LM = *L. mormyrus* and OM = *Oblada melanura*); reference source; forward and reverse primer sequences; and the optimised Ta (°C) used in the present study for each species (DC= *D. cervinus*, SP= *Spondyliosoma spp*., LM= *L. mormyrus* and SS= *S. salpa*).

ID	Species	Reference	Primers F	Primers R	DC Ta°C	SP Ta°C	LM Ta°C	SS Ta°C
DsaMS16	*DS*	Perez *et al*. (2008)	F: AGTCAAACCTCGGCATCAAGCGGGTA	R: ACGAGGAGCTCTGACTTCTGATTCGTT	55	-	-	-
DsaMS27	*DS*	Perez *et al*. (2008)	F: GCTCACTGTGCTGGCTCCACATCACC	R: GCGCTGTGCTTGCTGTCGGAGA	55	55	55	55
DsaMS34	*DS*	Perez *et al*. (2008)	F: AGATCAGATTTGCTGTGATAGCGTCCAAAG	R: ACTCCTGCAGCTCCTCCTGGGCTTC	55	55	55	55
DsaMS48	*DS*	Perez *et al*. (2008)	F: ACATCGCACACCCCCACAACC	R: TGCATGAACAACTTCCACACACAAGTCC	50	50	50	50
Dvul33	*DV*	Roques *et al* (2007a)	F: GCCGGGCTCGACATTGACACTGAA	R: GCAGCCAGCAGAGCTTAAAGAACT	50	50	50	50
Dvul38	*DV*	Roques *et al* (2007a)	F: TCGGGCACAGATAGAAAGAAACAC	R: GAAGGAAGACGGATCTCAGGATGA	50-55	47-51	47-57	47-51
Dvul4	*DV*	Roques *et al* (2007a)	F: GCGGTTATGTATACGTTGCGTTTA	R: TTGGCGTTGAACAGAAGTCAGACA	55	55	55	55
Dvul61	*DV*	Roques *et al* (2007a)	F: TGGGGACTCTCAGAATCATCACAA	R: TGGAAAAAGCCCTCTGGACAAAAG	-	55	57	47-57
Dvul84	*DV*	Roques *et al* (2007a)	F: GCTCGACGTGCACTCTGCCCTTGA	R: ATTCCCCAAATCCAGCACTCACAT	51-57	50-55	53-57	50-55
LM12	*LM*	Sala-Bozano *et al (*2009)	F: ACGGTATGGAGTCAACTGC	R: GAGTGTTTCTGACAGGATGAGAAC	50	50	50	50
LM19	*LM*	Sala-Bozano *et al (*2009)	F: AAACACACCTTCCCTCTCCT	R: AGCTGCTCAAACAGGCTATGA	-	-	47-57	47-53
LM68	*LM*	Sala-Bozano *et al (*2009)	F: CTTCAGGGGCGTTTCAGC	R: ACCAGGACAGGACCAGGTG	55	-	55	-
LM72	*LM*	Sala-Bozano *et al (*2009)	F: ATTGCAGACATGTCAGATGC	R: TATGCTCTAATGGCAATGCTC	-	-	50-55	-
LM86	*LM*	Sala-Bozano *et al (*2009)	F: ACTGCTGGCCTCTCTTTTGA	R: TGCCTCGTAATCCTCCACTT	50	50	50	50
Omel20	*OM*	Roques *et al.* (2007b)	F: CAGGGTAGCAACAGGGTAACAATG	R:GGCGGTTGAGGACACTGCAAAAAA	-	-	55	-
Omel38	*OM*	Roques *et al.* (2007b)	F: AGCCGGCTGAGCTCCATAATAACC	R: TGCCCTCTTGTCACACCAGGTCAC	55	55	55	55
Omel58	*OM*	Roques *et al.* (2007b)	F: GGCATTATTGTTCCATCATTACTCC	R: ATGGCATACAACCTGCATCAGAAG	55-56	50-55	47	48-56
Omel61	*OM*	Roques *et al.* (2007b)	F: CAGCGGGGGATTAATCTGCATTTG	R: GCCCGATTTATCTTCATCACCCAT	53-57	47-57	50	-

2.2.6 Microsatellite analysis

Numbers of alleles (N_A), allelic richness (A_R; El Mousadik and Petit, 1996), observed heterozygosity (H_O), and expected heterozygosity (H_E), were calculated using FSTAT 2.9.3.2 (Goudet, 1995). Genotype frequency conformance to Hardy-Weinberg equilibrium (HWE) expectations within loci and genotypic linkage equilibrium between pairs of loci were tested using exact tests in GENEPOP 3.3 (Raymond and Rousset, 1995). Genetic differences among samples were quantified by the unbiased F_{ST} estimator, θ (Weir and Cockerham, 1984), with significant differences from zero inferred following 10,000 permutations (Goudet *et al*., 1996) performed in FSTAT. The presence of Null alleles in the data set was explored using the programme FreeNA, where appropriate F_{ST} values corrected for null alleles were calculated in FreeNA with significance assessed after 1000 bootstraps.

Genetic structure was also investigated with the program STRUCTURE (Pritchard *et al*., 2000). STRUCTURE uses a Bayesian model-based clustering method to infer population structure assuming Hardy-Weinberg Equilibrium (HWE). STRUCTURE assumes a model where there are a number of populations (K) which may be unknown each of which are defined by a specific set of allele frequencies at each locus. Individuals in the samples are then assigned to a population, based on these criteria (Hubisz, *et al*., 2009). STUCTURE can be run using a number of models (and differing combinations thereof); the simplest being the 'no-admixture model', which assumes that each individual belongs to a single cluster or population. The more complicated 'admixture' model allows an individual to cluster with more than one population (Hubisz, *et al*., 2009). STRUCTURE may also be run with or without *a priori* sample partitioning. Using *a priori* sample partitioning allows STRUCTURE to place significantly more weight on clustering outcomes which are correlated with the sampling locations (Hubisz, *et al*., 2009), when the data suggest that this would be useful. Alternatively STRUCTURE may be run with no *a priori* sample partitioning. In the present study the analysis was run to identify the number of clusters, *K* (from a range of 1-5), with the highest posterior probability. Both the 'no admixture model' (as recommended for low F_{ST}; Pritchard *et al*., 2000) and 'admixture model with correlated allele frequencies' were employed. These analyses were repeated for both *a priori* and no *a priori* sample partitioning. Each MCMC run consisted of a burn in of 20000 steps followed by 80000 steps. Ten replicates were conducted for each *K* to assess

consistency. The *K* value best fitting the data set was estimated using both the log probability of data [Pr(X/K] and Evanno's delta K method (Evanno *et. al.*,2005) carried out in STRUCTURE HARVESTER (Earl and VonHoldt, 2012). STRUCTURE plots were prepared using CLUMPP (Jakobson and Rosenberg, 2007) and displayed using *distruct* (Rosenberg, 2004).

A major shortcoming of clustering methods such as STRUCTURE and *F*- statistics is the assumption of Hardy-Weinberg and linkage equilibrium within populations. However, often this assumption is breached. As such it is pertinent to utilise methods which do not require HWE or linkage equilibrium. In the present study Discriminant Analysis of Principal Components (DAPC) is employed to identify genetic clusters (Jombart *et al*., 2010). DAPC firstly transforms data using Principal Components Analysis (PCA) as a prior step to discriminant analysis (DA), which ensures that variables submitted to DA are perfectly uncorrelated. The DAPC function constructs synthetic variables, discriminant functions (DFs) that maximise variation between groups, whilst minimising variation within, groups and computes coordinates along these functions for each individual. DAPC can be performed either with or without group priors (Jombart *et al*., 2010). DAPC was implemented in the R-package *adegenet* (Jombart, 2008; R Core Team, 2017. To run the DAPC without priors the *find.clusters*() function was used for K = 1-5. The best supported number of clusters was estimated through comparison of the Bayesian Information Criterion (BIC) for the different values of K. The described relationships between the inferred clusters was explored using the *dapc()* function. To run the analysis with priors the analysis was run using the *dapc()* function with clusters inferred from the original sampling sites.

Assignment tests were undertaken assigning individuals to their respective regions (Angola or South Africa) using GeneClass 2 (Piry *et al*., 2004). Assignment of individuals was made using a Bayesian approach using the Rannala and Mountain (1997) algorithm which does not assume HWE. With a probability computation (Monte-Carlo resampling) using the Paetkau *et al*. (2004) algorithm with 10000 simulated individuals.

2.2.7 Power Analysis

It is crucial to assess the power of genetic data when using genetic markers to infer the spatial structuring of populations (Putman and Carbone, 2014). Failure to do so could lead to incorrect assumptions of population structuring. POWSIM 4.1 (Ryman and

Palm, 2006) was used to estimate the probability of Type I (α) error (a rejection of the null hypothesis of genetic homogeneity when it is true) and Type II (β) error (a failure to reject the null hypothesis of genetic homogeneity when it is false). POWSIM estimates α error rate and power as the proportion of random sub-samples that show statistically significant ($p < 0.05$) genetic differentiation after a base population, simulated from observed allele frequencies, undergoing genetic drift (with no migration or mutation) for a specified number of generations (*t*).

For both mtDNA markers Type I (α) error ($t = 0$) and Type II (β) error were estimated in POWSIM 4.1 using both chi-square and Fisher's exact tests with the following settings: 10 (COI) and 9 (CR) subpopulations, Ne= 2000, 1000 iterations and sample sizes were adjusted as recommended by Larsson *et al*. (2009) to reflect the haploid nature of mtDNA. *t* was adjusted to test for the following F_{ST} values 0.005- 0.07 (i.e. until power to resolve F_{ST} value ≥95%). Whilst for the microsatellite data set power was estimated again using both chi-square and Fisher's exact tests under the following conditions: 4 loci, 4 subpopulations, Ne= 2000, no modification of sample size (since nuclear microsatellites are diploid) and 1000 iterations. The power to resolve F_{ST} values from 0.001- 0.01 by the microsatellite data set was assessed as well as for F_{ST}= 0 (α error).

2.3 Results

2.3.1 Mitochondrial DNA diversity

Sequences for a 550bp COI fragment from 118 individuals are included in this study: 74 individuals from the UK, Angola and South Africa (see Table 2.1) sequenced in the present study, plus 44 sequences obtained from GenBank representing *Spondyliosoma* from Turkey (21), Portugal (16) and South Africa (7). In total, these sequences yielded 39 polymorphic sites which defined 22 COI haplotypes. Overall genetic diversity was moderately high (h= 0.5202 (SD= 0.0889), π = 0.001831 (SD= 0.001386)). Regionally, genetic diversity was found to be highest in Turkey (h= 0.7571 (SD = 0.0646), π = 0.003429 (SD=0.002275)), followed by the North East Atlantic (h= 0.5785 (SD = 0.1004), π = 0.001916 (SD = 0.001453)), South Africa (h = 0.5573 (SD = 0.1089), π = 0.001624 (SD = 0.001310)) and considerably lower in Angola (h = 0.1878 (SD = 0.0815), π =0.000355 (SD = 0.000505)) – see Table 2.3. Genetic diversity for individual sampling sites is reported in Table 2.4, which show similarly low h and π values across Angolan sites but differences between sites in South Africa and the North East Atlantic / Mediterranean (Portugal more similar to Turkey, with Wales, UK having substantially lower diversity).

For the CR, 87 individuals were sequenced for 436bp of the hypervariable region I, yielding 76 polymorphic sites and 34 haplotypes in total. Overall genetic diversity for CR in *Spondyliosoma* was moderately high with h = 0.7031 (SD= 0.0888) and π = 0.004369 (SD = 0.00386). The observed pattern of CR genetic diversity was similar to that found in COI with highest genetic diversity in the North East Atlantic (h= 0.7917 (SD = 0.0886), π = 0.005810 (SD = 0.003692)) followed by South Africa (h= 0.7692 (SD = 0.0689), π = 0.005516 (SD = 0.003399)) and lowest in Angola (h= 0.5484 (SD = 0.1090), π = 0.001780 (SD = 0.001486)) – see Table 2.5. Similar variability in CR genetic diversity indices among individual sampling sites was observed (Table 2.6).

Table 2.3 Genetic diversity for *Spondyliosoma spp.* mtDNA COI sequences for identified clades/ regions. n: sample size; H: haplotype number; PH private haplotype number; π: nucleotide diversity; h: haplotype diversity; τ: tau; PSSD: the probability that the empirical distribution of mismatches was significantly different from the distribution simulated under a demographic expansion model; Texp= time since expansion (Kya); D: Tajima's D; and F_s =Fu's FS. Bold indicates significant values. * denotes region contains samples from GenBank

	Angola	**South Africa**	**NE Atlantic***	**Turkey***
N	41	18	31	21
H	5	5	7	5
PH	5	5	7	5
h (SD)	0.188 (0.082)	0.557 (0.109)	0.579 (0.100)	0.757 (0.065)
π(SD)	0.0004 (0.0005)	0.0016 (0.0013)	0.0019 (0.0015)	0.0034 (0.0023)
τ (95% CI)	3.000 (0.359, 3.500)	1.393 (0.000, 3.746)	2.164 (0.000, 4.824)	2.451 (0.545, 4.561)
PSSD (P)	0.0011 (0.407)	0.0011 (0.923)	0.0036 (0.840)	**0.1533 (0.023)**
Raggedness (P)	0.4291 (0.663)	0.0518 (0.924)	0.0514 (0.894)	**0.5530 (0.002)**
Texp (95% CI)	227 Kya (27/ 265)	106 Kya (Present Day/ 284)	164 Kya (Present Day/ 366)	-
D(P)	**-1.878 (0.004)**	-0.487 (0.354)	-0.845 (0.217)	0.411 (0.693)
F_S(P)	**-5.093 (0.000)**	-1.230 (0.145)	**-2.499 (0.047)**	0.603 (0.649)

Table 2.4 Genetic diversity of sampling sites for *Spondyliosoma spp*. using mtDNA COI sequences. n: sample size; H: haplotype number; PH private haplotype number; h: haplotype diversity; π: nucleotide diversity. Site ID codes can be found in Table 2.1. * denotes sequences from GenBank.

	BEN	**LUC**	**NAM**	**TOM**	**FAL**	**MOB**	**ABE**	**POR***	**TUR***
N	13	8	11	9	9	9	15	16	21
H	2	2	2	2	3	2	5	6	5
PH	1	1	1	1	2	1	1	2	5
h (SD)	0.154 (0.126)	0.250 (0.180)	0.182 (0.144)	0.222 (0.166)	0.417 (0.191)	0.222 (0.166)	0.476 (0.155)	0.683 (0.120)	0.757 (0.065)
π(SD)	0.0003 (0.0005)	0.0005 (0.0007)	0.0003 (0.0005)	0.0004 (0.0006)	0.0008 (0.0009)	0.0012 (0.0012)	0.0014 (0.0012)	0.0024 (0.0018)	0.0034 (0.0023)

Table 2.5 Genetic diversity per region / clade for *Spondyliosoma spp.* mtDNA CR sequences. n: sample size; H: haplotype number; PH private haplotype number; π: nucleotide diversity; h: haplotype diversity; τ: tau; PSSD: the probability that the empirical distribution of mismatches was significantly different than the distribution simulated under a demographic expansion model; Harpending's Raggedness Index; T exp= Time since expansion (Kya); D: Tajimas D; and F_s =Fu's FS; (P) = associated p- value. **Bold** indicates significant values.

	Angola	**South Africa**	**NE Atlantic**
N	31	40	16
H	11	16	7
PH	11	16	7
h (SD)	0.548 (0.109)	0.769 (0.069)	0.792 (0.089)
π(SD)	0.0018 (0.0015)	0.0055 (0.0034)	0.0058 (0.0037)
τ (95% CI)	0.801 (0.287, 1.557)	1.322 (0.139, 2.371)	1.125 (0.000, 13.068)
PSSD (P)	0.0001 (0.942)	0.0056 (0.322)	0.0326 (0.392)
Raggedness (P)	0.0671 (0.811)	0.0698 (0.346)	0.1235 (0.267)
T exp (95% CI)	26 Kya (9 / 50)	42 Kya (4 / 76)	36 Kya (Ongoing / 416).
D (P)	**-2.405 (0.000)**	**-2.312 (0.001)**	**-1.554 (0.047)**
F_S (P)	**-10.408 (0.000)**	**-8.073 (0.001)**	-1.012(0.268)

Table 2.6 Genetic diversity per sampling site for *Spondyliosoma* spp. mtDNA CR sequences. n: sample size; H: haplotype number; PH private haplotype number; h: haplotype diversity; π: nucleotide diversity. Sample ID codes are as per Table 2.1.

	BEN	**LUC**	**NAM**	**TOM**	**FAL**	**MOB**	**TSI**	**POR**	**ABE**
N	9	5	12	5	13	15	6	6	16
H	5	3	3	3	8	8	2	4	7
PH	4	0	2	2	5	5	0	0	7
h(SD)	0.722 (0.159)	0.700 (0.218)	0.318 (0.164)	0.700 (0.218)	0.808 (0.113)	0.838 (0.085)	0.333 (0.215)	0.867 (0.129)	0.792 (0.087)
π(SD)	0.0026 (0.0021)	0.0028 (0.0024)	0.0008 (0.0009)	0.0018 (0.0018)	0.0063 (0.0040)	0.0072 (0.0044)	0.0008 (0.0010)	0.0043 (0.0033)	0.0058 (0.0037)

2.3.2 Phylogenetics

COI phylogeny inference by median joining network identified four highly divergent and geographically distinct groups of haplotypes (clades) corresponding to individuals from South Africa, Angola, North East Atlantic and Turkey (Figure 2.2). The Turkey clade appears to be the most divergent, with the remaining three 'Atlantic' clades (North East Atlantic, Angola and South Africa) being roughly equally divergent from one another. Within clades it is possible to identify common haplotypes, with Angola in particular being dominated by a single common haplotype plus four singleton haplotypes, whereas the Turkey clade does not exhibit any clear common haplotype (Figure 2.3).

Optimal models of nucleotide variation for the COI sequence data set carried out in jModelTest 2.1.4 identified the HKY + G (0.1450) model as best fitting and was used when constructing phylogenies. The Maximum Likelihood tree resolved the same four reciprocally monophyletic and geographically consistent clades with high statistical support (> 85%) and identified the Turkish clade as the most divergent (Figure 2.4). With the inclusion of other Sparid species as outgroups the South African clade is identified as being basal, with the next derived clade being the North East Atlantic, then the Angolan and Turkey clades (Figure 2.4).

Net COI divergence (Da) as calculated in MEGA 6 identified sequence divergence between the four clades ranging from 2.0 to 3.1% (Table 2.7). The largest estimated net divergence was between the South African and Mediterranean clades at 3.1 % (SE 0.7%), yielding an estimated divergence (assuming a standard 1.2%/Myr COI sequence divergence) time of 2.58 Ma (SE 2.00- 3.17 Ma). Da between the North East Atlantic and Mediterranean clade is 3.0 % (SE 0.7%) giving an estimated date of divergence of 2.50 Ma (SE 1.92-3.08 Ma). The Da between Angola and the Mediterranean was 2.7% (SE 0.6%), with an estimated date of divergence of 2.25 Ma (SE 1.75 -2.75 Ma). The three Atlantic clades have similar levels of divergence from each other and are less divergent from one another than each is from the Mediterranean clade. The South African and Angolan clades exhibited a Da of 2.3% (SE 0.6%) yielding an estimated divergence time of 1.92 Ma (SE 1.42- 2.42 Ma). The South African and North East Atlantic clades yielded a Da of 2.1 % (SE 0.6%), with an estimated date of divergence of 1.75 Ma (SE 1.25-2.25 Ma). Finally the Angolan and North East Atlantic clades exhibited the smallest observed Da of 2% (SE 0.6%) and an estimated date of divergence of 1.67 Ma (SE 1.17-2.17 Ma). Mean intraclade divergence was much smaller than inter-clade divergence, the highest being amongst Turkey haplotypes (0.34%) and lowest in Angola (0.04%) – see Table 2.7.

Table 2.7 Estimates of % sequence divergence (below diagonal: Da ±SE) and times since divergence (above diagonal: Ma ±SE) in *Spondyliosoma* spp. based upon the COI data set. Values on diagonal represent % intra-clade mean sequence divergence.

	Angola	**South Africa**	**NE Atlantic**	**Turkey**
Angola	0.04% (0.017 %)	1.92 Ma (1.42- 2.42)	1.67 Ma (1.17- 2.17).	2.25 Ma (1.75-2.75)
South Africa	2.30% (0.6%)	0.15% (0.093%)	1.75 Ma (1.25-2.25).	2.58 Ma (2.00 - 3.17).
NE Atlantic	2.00% (0.6%)	2.10 % (0.6%)	0.19% (0.09%)	2.50 Ma (1.92- 3.08).
Turkey	2.70% (0.6%)	3.10 % (0.7%),	3.00 % (0.7%)	0.34% (0.14%)

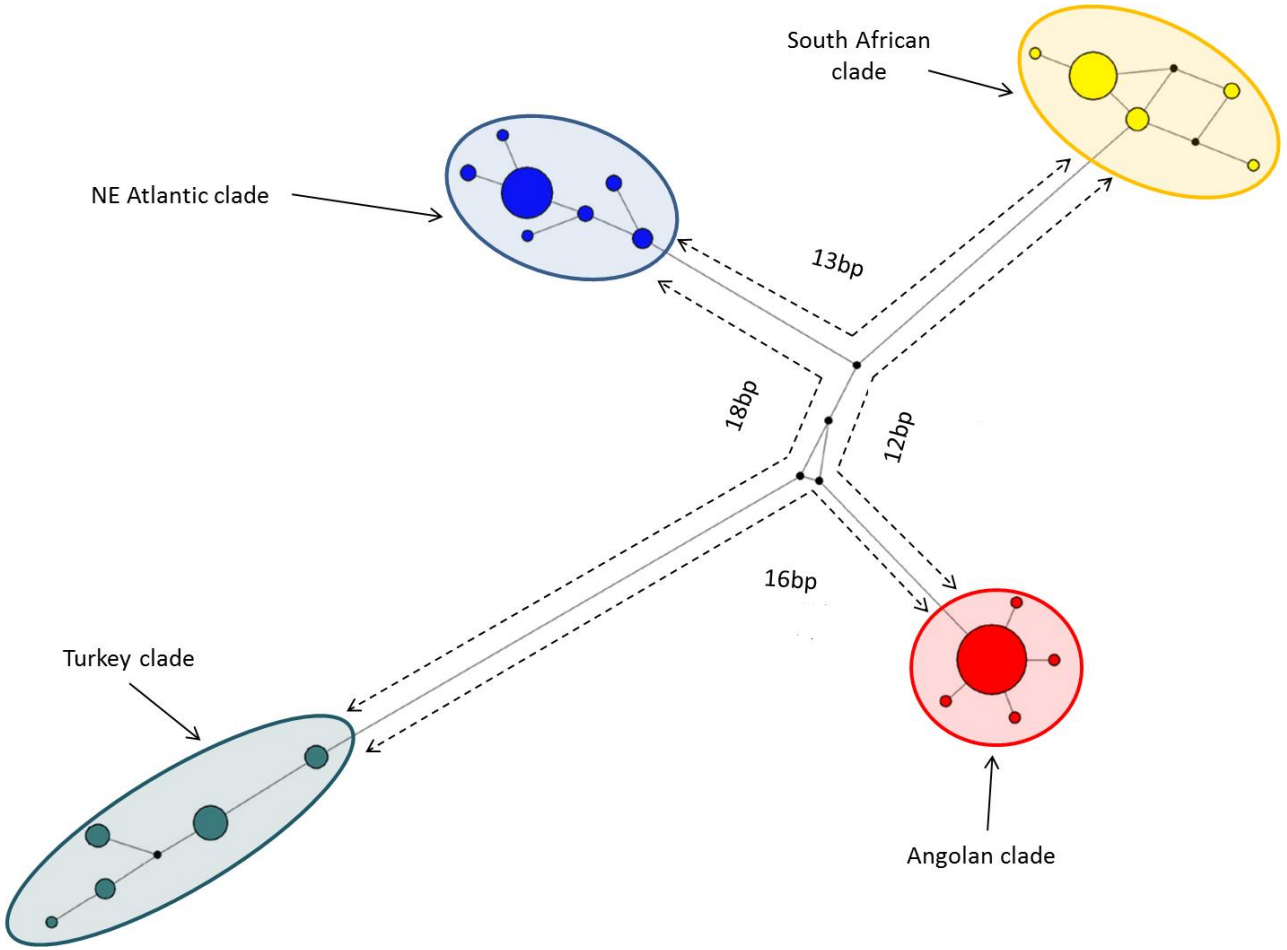

Figure 2.3 Reconstructed median-joining haplotype network for *Spondyliosoma* spp. using 550bp of mtDNA COI. Node sizes are proportional to the observed number of individuals bearing that haplotype, with the smallest nodes representing a single individual. Predicted missing (or extinct) haplotypes are represented by small black nodes. Colouring refers to geographical regions, from which individuals were sampled, with Yellow corresponding to South Africa, Red to Angola, Blue to NE Atlantic and Grey to Eastern Mediterranean. Defined clades are encircled.

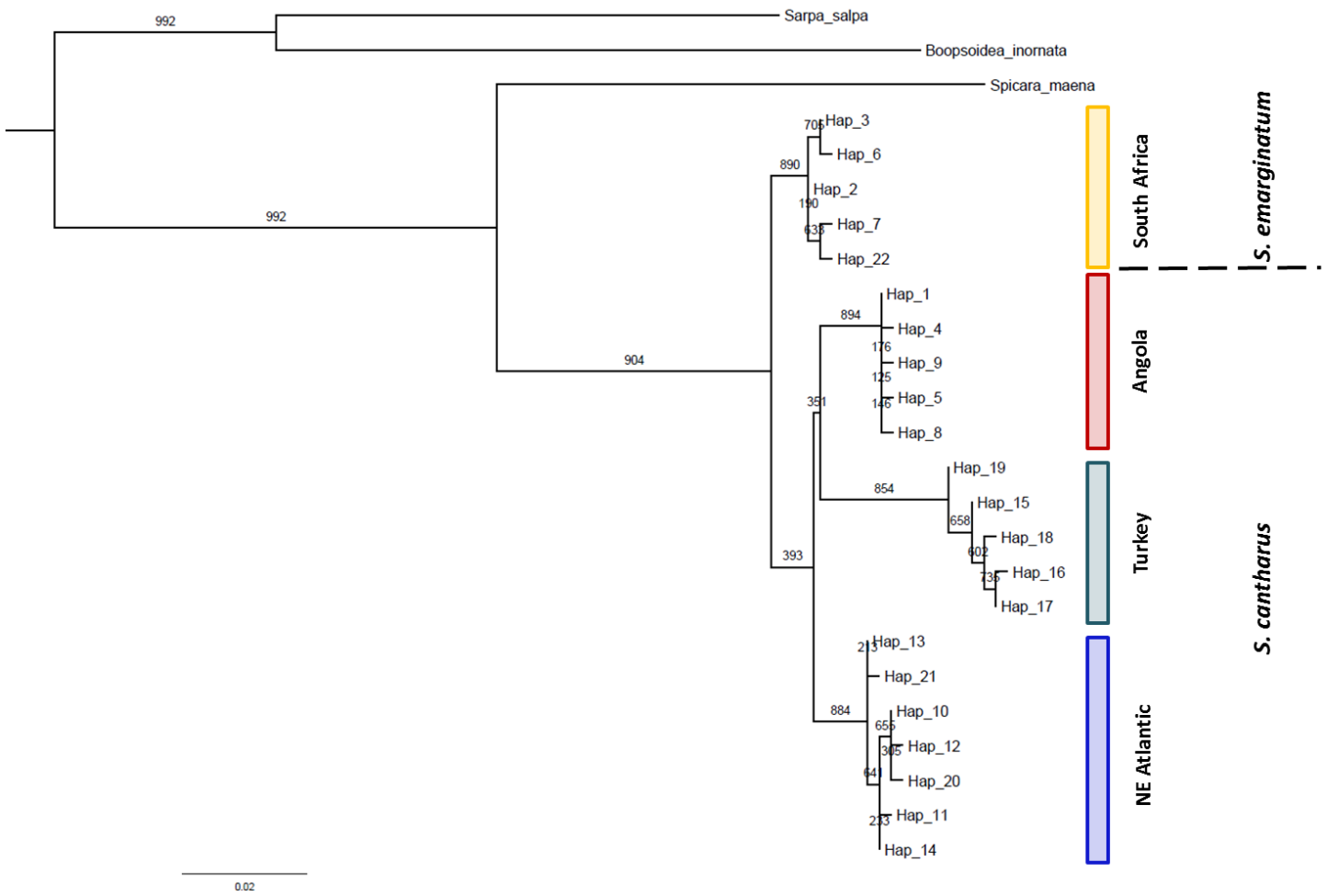

Figure 2.4 Maximum likelihood reconstruction of phylogenetic relationships among mtDNA COI haplotypes in *Spondyliosoma spp.* Bootstrap values are displayed on the branch (1000 bootstraps). The four major clades defined by the present study, as well as the recognised split between the two described species of *Spondyliosoma*, are highlighted.

Investigations into the relationships amongst the identified CR haplotypes utilising a Median joining algorithm in NETWORK also recovered the three Atlantic clades identified utilising the COI dataset (Figure 2.5). The South African clade occupies a central position between the North East Atlantic and Angolan clades (Figure 2.5). Both Angolan and South African clades exhibited a single common haplotype (Angolan common haplotype found in 67% of individuals; South African common haplotype found in 48% of individuals). Angola exhibits one common haplotype and ten singleton haplotypes all differing by a single base pair from the common haplotype, whilst South Africa exhibits a similar pattern but with several singletons which are considerably divergent from the common haplotype (Figure 2.5). The North East Atlantic exhibits fewer haplotypes overall compared to the other two clades and there is no obvious common haplotype.

jModelTest 2.1.4 identified the TPM2UF + I (0.6230) as being the best fitting model for nucleotide variation in the CR dataset. However this model is unavailable in PhyML, and as such the next best fitting model of HKY + I (0.6220) was utilised in phylogenetic reconstruction. The tree identified the same three reciprocally monophyletic clades with high support (Figure 2.6), and identified the North East Atlantic clade as being basal with Angola and South Africa being sister clades (Figure 2.6). Net divergence estimates identified the North East Atlantic clade to be the most divergent (Table 2.8), with Da between the South African and North East Atlantic clades at 8.6% (SE 1.3 %), giving an estimated divergence time of 2.39 Ma (SE 2.03-2.75 Ma), and Da between Angolan and North East Atlantic clades at 8.2% (SE 1.2%), giving an estimated divergence time of 2.28 Ma (SE 1.94- 2.61 Ma). Finally the Angolan and South African clades yield a Da of 7.7% (SE 1.3%) with an estimated date of divergence of 2.13 Ma (SE 1.78-2.5 Ma). Estimated intraclade divergence was much lower than between-clade divergence, being highest within the North East Atlantic clade (0.58%) and lowest in Angolan clade (0.12%) - Table 2.8.

To further investigate the relationship of the three Atlantic *Spondyliosoma* clades COI and CR sequences were concatenated for further phylogenetic analyses. Concatenation of the COI and CR was possible for 48 individuals yielding a combined sequence length of 991bp, 24 haplotypes and 90 polymorphic sites. Reconstruction of the haplotype network revealed the three previously identified clades corresponding geographically to North East Atlantic, Angolan and South African samples (Figure 2.7). Likewise the ML

phylogeny identified and supported the three previously identified North East Atlantic, Angolan and South African clades, with the South African clade identified as being basal to the other two clades (Figure 2.8).

Table 2.8 Estimates of % sequence divergence (below diagonal: Da +SE) and times since divergence (above diagonal: Ma +SE) in *Spondyliosoma* spp. based upon the CR data set. Values on diagonal represent % intra-clade mean sequence divergence.

	Angola	**South Africa**	**NE Atlantic**
Angola	0.18% (0.05%)	2.13 Ma (1.78-2.50).	2.28 Ma (1.94- 2.61)
South Africa	7.70% (1.30%)	0.55% (0.11%)	2.39 Ma (2.03-2.75).
NE Atlantic	8.20% (1.20%)	8.60% (1.30 %)	0.58% (0.18%)

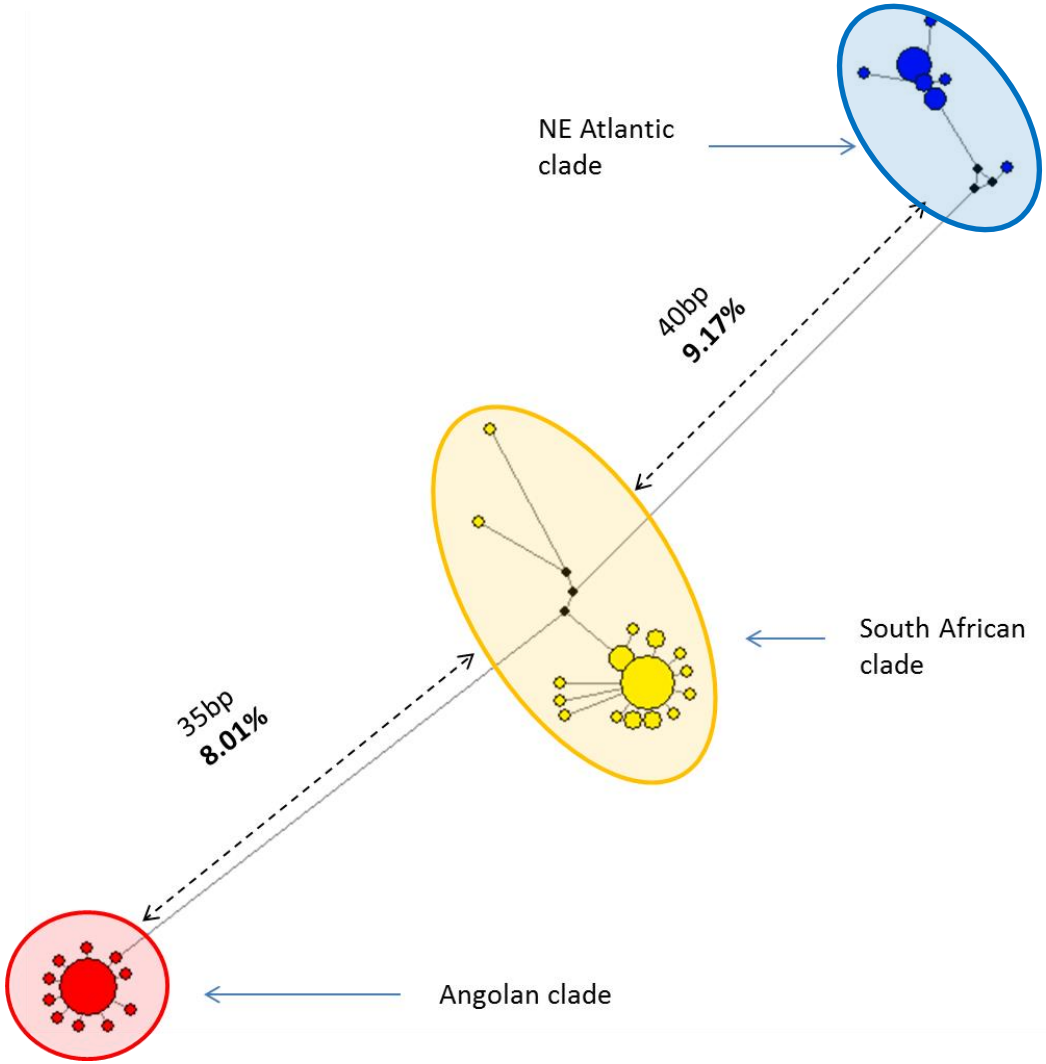

Figure 2.5 Reconstructed median-joining haplotype network for *Spondyliosoma* based on mtDNA CR. Node sizes are proportional to the observed number of individuals bearing that haplotype. Colouring refers to geographical regions, from which individuals were sampled, with Yellow corresponding to South Africa, Red to Angola and Blue to NE Atlantic. Defined clades encircled.

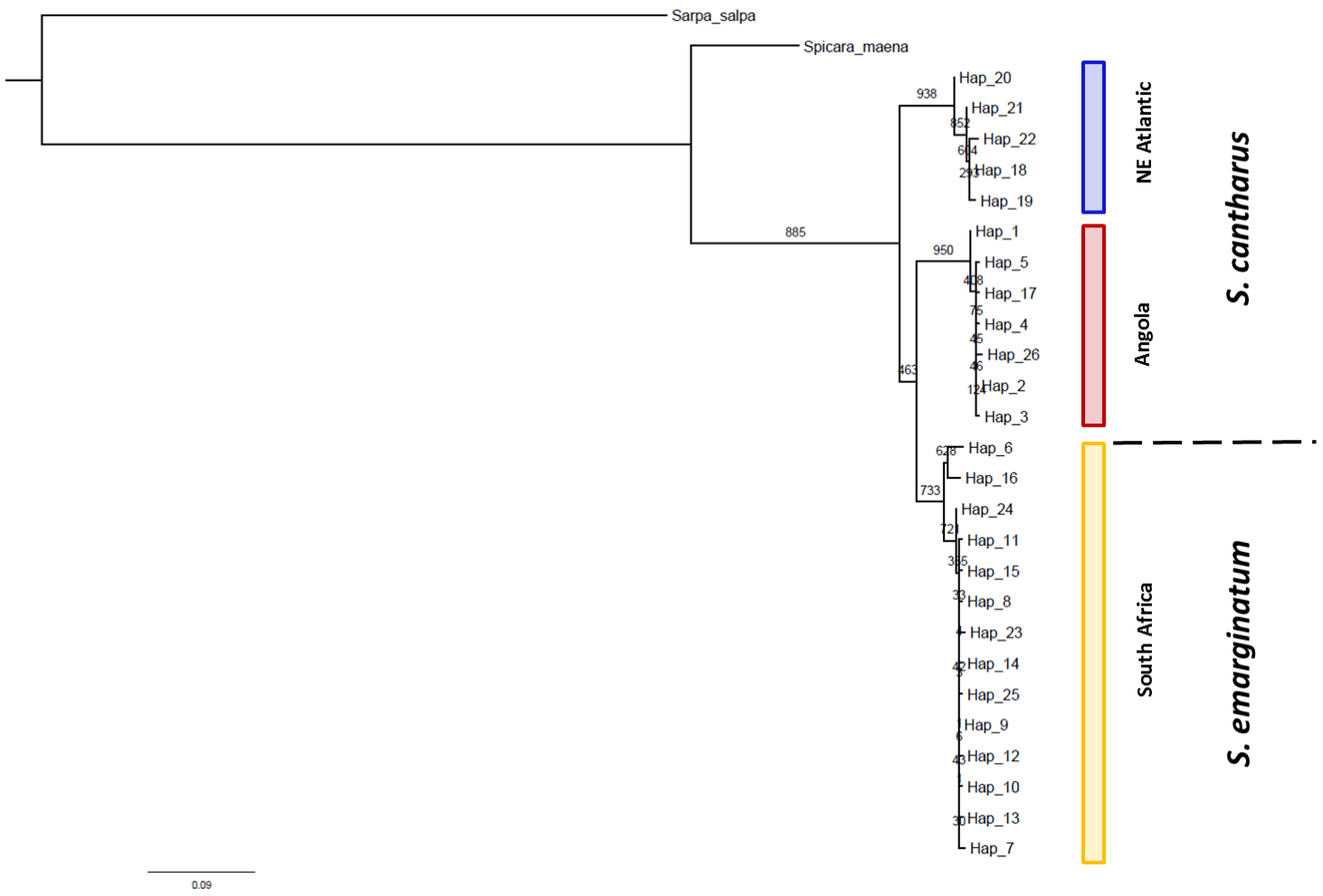

Figure 2.6 maximum likelihood reconstructions of phylogenetic relationships among mtDNA CR haplotypes in *Spondyliosoma*. Bootstrap values are displayed on the branch (1000 bootstraps). The three major clades defined by this study, as well as the recognised split between the two described species of *Spondyliosoma*, are highlighted.

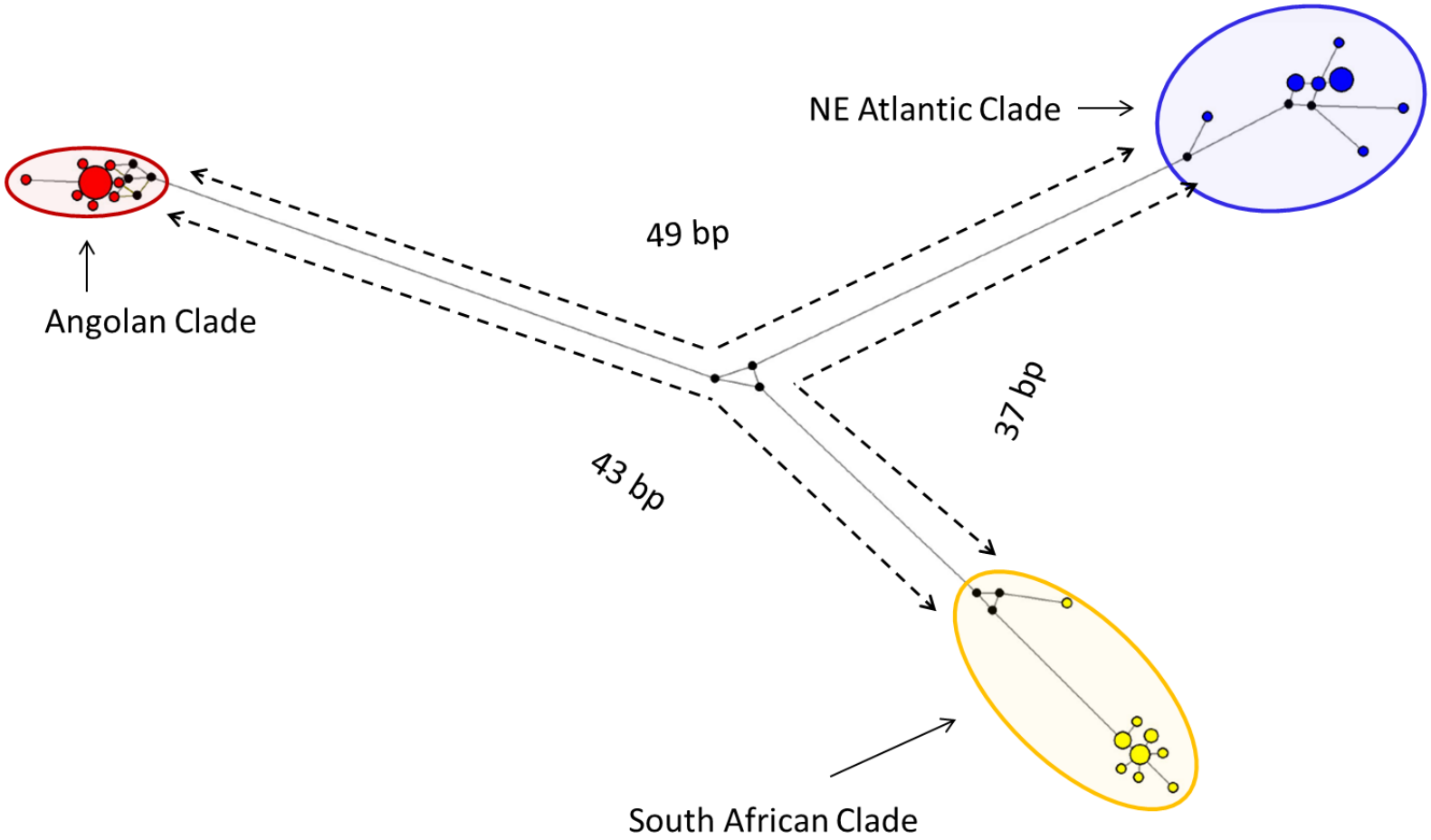

Figure 2.7 Reconstructed median-joining haplotype network for *Spondyliosoma spp*. based on concatenated mtDNA COI and CR sequences. Node sizes are proportional to the observed number of individuals bearing that haplotype, with the smallest node corresponding to a single individual. The identified circled clades are the same as those identified using the CR sequences alone and are labelled as such.

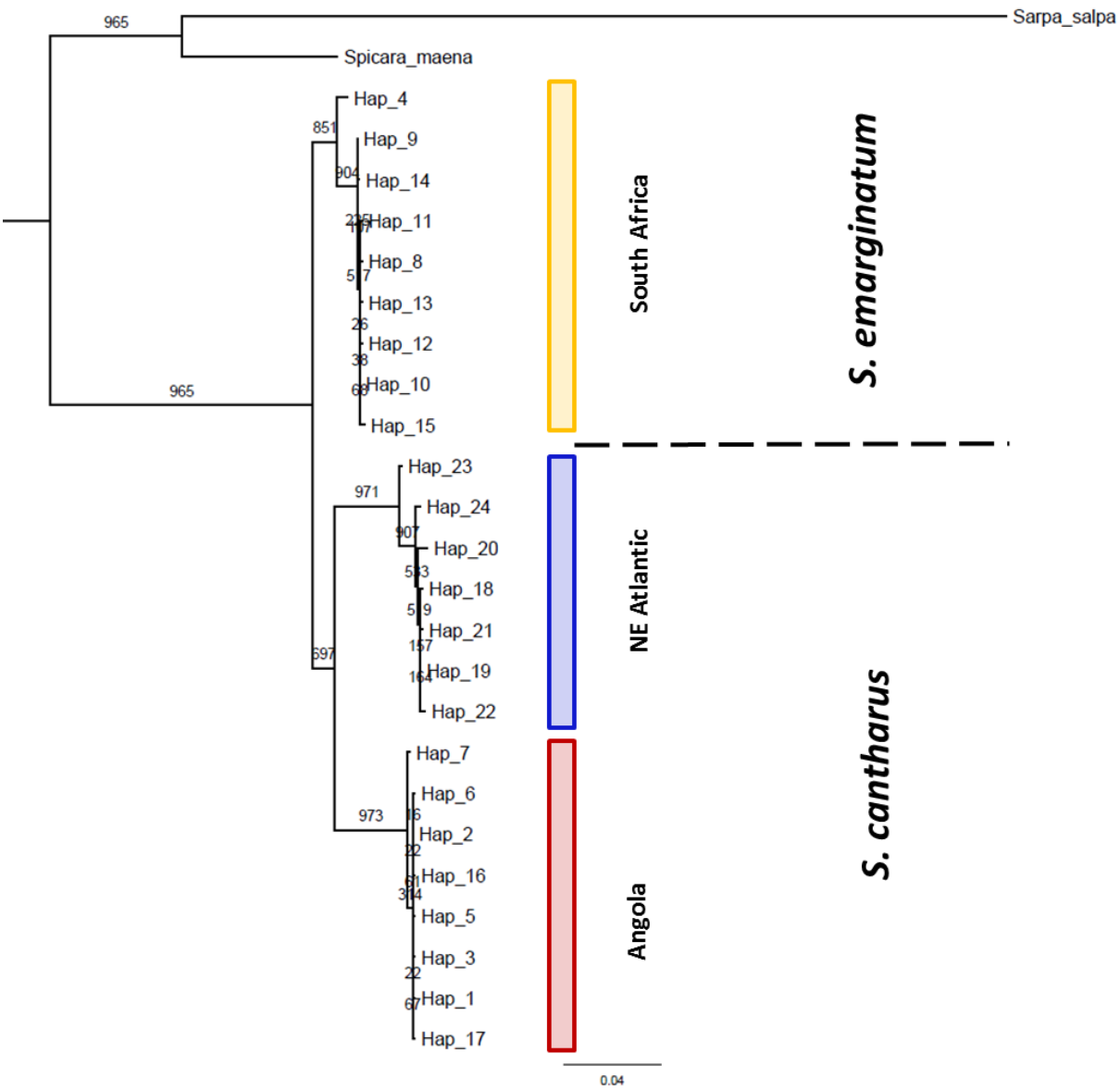

Figure 2.8 maximum likelihood reconstructions of phylogenetic relationships among the observed haplotypes from the concatenated mtDNA COI-CR sequences in *Spondyliosoma spp.* Bootstrap values are displayed on the branch (1000 bootstraps). The three major clades defined by this study, as well as the recognised split between the two described species of *Spondyliosoma*, are highlighted.

2.3.3 mtDNA population structuring

For both COI and CR data sets pairwise Φ_{ST} values between *Spondyliosoma* samples within Angola and South African were all low (COI = 0.000-0.007; CR = 0.000-0.087) and not significantly different from zero, whilst the Φ_{ST} values between samples from the two regions were all high (COI = 0.914-0.979; CR = 0.926-0.991) and significant (Tables 2.9 and 2.10). Pairwise Φ_{ST} comparisons between all four geographical regions were in excess of 0.92 and significant (Tables 2.11 and 2.12). The identified high level of phylogeographic structure translated into large and significant among-region variance (COI = 96.38%; CR = 95.32%) but within-region variance was much smaller (COI = 0.81%; CR = -0.1%) in the AMOVA (Table 2.13)

Table 2.9 Pairwise Φ_{ST} values between *Spondyliosoma* spp. samples based upon mtDNA COI sequence data. Sample codes as per Table 2.1. Bold indicates values significantly different from zero.

	BEN	**LUC**	**NAM**	**TOM**	**FAL**	**MOB**	**ABE**	**POR**
LUC	0.013	-						
NAM	0.001	0.006	-					
TOM	0.007	0.001	0.002	-				
FAL	**0.979**	**0.973**	**0.977**	**0.975**	-			
MOB	**0.973**	**0.964**	**0.970**	**0.966**	-0.047	-		
ABE	**0.961**	**0.958**	**0.958**	**0.954**	**0.951**	**0.944**	-	
POR	**0.931**	**0.926**	**0.926**	**0.920**	**0.920**	**0.913**	0.002	-
TUR	**0.924**	**0.920**	**0.920**	**0.914**	**0.921**	**0.918**	**0.921**	**0.908**

Table 2.10 Pairwise Φ_{ST} values between *Spondyliosoma spp*. samples based upon mtDNA CR sequence data. Sample codes as per Table 2.1. Bold indicates values significantly different from zero

	BEN	**LUC**	**NAM**	**TOM**	**FAL**	**MOB**	**TSI**	**PEL**
LUC	0.004	-						
NAM	0.017	0.087	-					
TOM	-0.015	0.060	0.06	-				
FAL	**0.942**	**0.935**	**0.956**	**0.939**	-			
MOB	**0.933**	**0.926**	**0.947**	**0.929**	-0.024	-		
TSI	**0.977**	**0.980**	**0.991**	**0.985**	-0.081	-0.063	-	
PEL	**0.960**	**0.956**	**0.977**	**0.961**	0.005	0.025	0.057	-
ABE	**0.947**	**0.943**	**0.958**	**0.944**	**0.934**	**0.929**	**0.951**	**0.941**

Table 2.11 Pairwise Φ_{ST} values for regions / clades based upon *Spondyliosoma* spp mtDNA COI sequence data. Bold indicates statistically significant Φ_{ST} values.

	Angola	**South Africa**	**NE Atlantic**
South Africa	**0.966**	-	
NE Atlantic	**0.923**	**0.952**	-
Turkey	**0.926**	**0.952**	**0.922**

Table 2.12 Pairwise Φ_{ST} values for regions / clades based upon the mtDNA CR sequence data, between *Spondyliosoma spp.* Bold indicates statistically significant values.

	Angola	**South Africa**
South Africa	**0.953**	
NE Atlantic	**0.939**	**0.964**

Table 2.13 Analysis of molecular variance results based on *Spondyliosoma* spp. COI and CR sequence data. The AMOVA was structured using Angolan and South African sample regions as groups.

Source of variation	Percentage of variation COI (p-value)	Percentage of variation CR (p-value)
Among groups	**97.73 (0.000)**	**95.32 (0.000)**
Among populations within groups	-0.05 (0.675)	-0.10 (0.689)
Within populations	2.32 (0.067)	**4.78 (0.032)**

2.3.4 Demographic tests

Given the pronounced inter-regional structuring and absence of any intra-regional structuring (where testable) demographic analyses for mtDNA were performed at the regional level pooling samples within regions.

For the South African sample, Fu's *Fs* and Tajima's *D* were negative but only significant for the CR sequence data set, indicating possible demographic expansion in South African *Spondyliosoma* (Tables 2.3 and 2.5). Mismatch distribution analysis could not exclude (i.e. was not significant) a hypothesis of population expansion in South African *Spondyliosoma*. The COI data set yielded $\tau = 1.39$ (95% CI 0.0 - 3.75), yielding an estimated time since expansion of ~106 Kya (95% CI ongoing -284 Kya). Using the CR data set yielded $\tau = 1.32$ (95% CI 0.14 - 2.37) giving estimated time since expansion of ~42 Kya (95% CI 4 -76 Kya).

For the Angolan sample Fu's *Fs* and Tajima's *D* were negative and significant for both COI and CR data, and mismatch distribution analysis could not exclude a hypothesis of population expansion (Tables 2.3 and 2.5). The COI data set yielded $\tau = 3.00$ (95% CI 0.36 - 3.50) giving an estimated time since expansion of ~227 Kya (95% CI 27 -265 Kya), whilst the CR data set gave $\tau = 0.80$ (95% CI 0.29 - 1.56) giving an estimated time since expansion of ~26 Kya (95% CI 9 -50 Kya).

For the North East Atlantic sample only Fu's *Fs* for the COI data and Tajima's *D* for the CR sequence data yielded a significant negative value (Tables 2.3 and 2.5), and mismatch distribution analysis could not exclude a hypothesis of population expansion. The COI data set yielded $\tau = 2.45$ (95% CI 0.55 - 4.56) giving an estimated time since expansion of ~164 Kya (95% CI ongoing -366 Kya), whilst the CR dataset gave a value of $\tau = 1.13$ (95% CI 0.00 - 13.07) giving an estimated time since expansion of ~36 Kya (95% CI ongoing -416 Kya).

For the Turkey sample demographic analyses could only be carried out on the COI data set due to the lack of published CR sequences. Both tests for neutrality yielded non-significant values, whilst mismatch distribution analyses did not support population expansion (Table 2.3).

2.3.5 Nuclear Microsatellite DNA diversity and structuring

Of the seven microsatellite loci tested only four (DsaMS27, DsaMS34, DsaMS48 and Dvul84) proved to be polymorphic after initial genotyping trials. Individuals from four sampling sites, two from South Africa (False Bay n = 27; Mossel Bay n = 21) and two from Angola (Benguela n = 27; Namibe n = 29), were genotyped for the four polymorphic loci. There were no significant deviations from random associations of genotypes (linkage disequilibrium) detected for any pair of loci, either across all samples (data pooled) or in any single sample, indicating that all loci are independent. Information on genetic variation for each sample / locus combination is provided in Table 2.14. All loci were variable in each sample with the total number of alleles per locus ranging from 18 (DsaMS48) to 28 (DsaMS34) with an average of 21.75 alleles. Whilst levels of variability differed across loci, multi-locus variability indices were similar across all samples (Table 2.14). All four loci exhibited significant global deviations from HWE. Significant deviations from HWE were found in 14 out of 16 locus/sample comparisons with only DsaMS27 and Dvul84 in Mossel Bay conforming to HWE. In all cases H-W departures were due to heterozygote deficits (Table 2.14). FreeNA identified the presence of null alleles in all but two of the locus/ sample comparisons (DsaMS27 / Mossel Bay and Dvul84 / Namibe) giving a likely explanation to the deviations from HWE found above.

Table 2.14 Genetic diversity of four microsatellite loci in *Spondyliosoma.* samples. N - number of individuals genotyped; Na - number of alleles; AR - allelic richness; H_E - expected heterozygosity; Ho - observed heterozygosity. pHWE- Hardy–Weinberg equilibrium probability. For sample Id codes see Table 2.1. Bold indicates statistically significant values.

		BEN	NAM	FAL	MOB
DsaMS27	**N**	25	28	25	21
	N_a	13	13	15	16
	AR	10.282	10.082	12.165	13.523
	H_E	0.854	0.858	0.896	0.908
	H_O	0.640	0.643	0.800	0.952
	pHWE	**0.001**	**0.004**	**0.043**	0.450
DsaMS34	**N**	25	26	21	18
	N_a	19	20	13	12
	AR	15.270	14.998	10.737	10.918
	H_E	0.920	0.917	0.810	0.832
	H_O	0.760	0.692	0.524	0.389
	pHWE	**0.017**	**0.000**	**0.039**	**0.000**
DsaMS48	**N**	19	24	20	14
	N_a	11	11	14	9
	AR	10.153	9.780	11.782	9.000
	H_E	0.861	0.880	0.884	0.814
	H_O	0.579	0.417	0.650	0.571
	pHWE	**0.004**	**0.000**	**0.001**	**0.005**
Dvul84	**N**	23	27	27	18
	N_a	4	8	13	10
	AR	3.962	5.545	10.312	9.548
	H_E	0.529	0.570	0.818	0.836
	H_O	0.348	0.556	0.630	0.667
	pHWE	**0.010**	**0.049**	**0.004**	0.245

Bayesian clustering supported a model of *K*=2 (Figure 2.9), with a clear geographical pattern (Angola and South Africa) of individual assignment evident in the STRUCTURE plots. Using the locprior model in STRUCTURE all individuals clustered into two groups according to their sampling region. Using the no locprior model some South African individuals (see Figure 2.9) showed genetic similarity to Angolan individuals. However upon performing classical assignment tests in GeneClass 2 these individuals assigned to their geographical sample origin. More broadly the assignment tests carried out in GeneClass 2 few individuals had a probability >0.95 in being assigned to their original sampling region, whilst some appeared to be equally assigned to both regions these individuals had missing loci (Table 2.15). The pattern of genetic structuring between Angolan and South African samples was also supported by

pairwise F_{ST} carried out in FSTAT, with all values between Angola and South Africa being high (F_{ST}> 0.09) and significantly different from zero while values between samples within regions were low and not significant (Table 2.16). Given the potential presence of null alleles at loci, F_{ST} values were also calculated using FreeNA with and without correcting for null alleles; however the significance of these values could not be estimated as significance by bootstrapping requires a minimum of five loci. F_{ST} values were high (>0.08) for comparisons between Angolan and South African samples both before and after correction for null alleles, albeit with F_{ST} values slightly lower after correction (i.e. null alleles may be upwardly biasing uncorrected F_{ST} estimates Table 2.17).

DAPC was run firstly with priors (defined as per sampling group) and secondly without priors using the *find.clusters* function in *adegenet*. The DAPC ran with priors retained the first 40 PC's representing 93.6% of variability. Once plotted the with priors DAPC identified no overlap of Angolan and South African individuals (Figure 2.10), with the first principle component discerning the clear difference between Angolan and South African individuals. The second principle component identified some differentiation between the two South African sampling sites, although the meaningfulness of such differentiation is questionable given the greatly reduced explanatory power of the second principle component (Figure 2.10). The DAPC run without priors using the *find.clusters* function resolved a K= 2, correctly assigning 86.54% of individuals (90 individuals) geographically to either South Africa or Angola. Of the misassigned individuals all 14 were geographically sampled in Angola but assigned by the *find.clusters* function to South Africa. For the subsequent DAPC 50 Principal Components were retained representing 97.34% of variance, once plotted the inferred clusters show little overlap (Figure 2.10). The Angolan inferred cluster has a less dense grouping of individuals than the South African cluster, suggesting that these individuals were misassigned by 'overlapping with the inferred South African cluster (Figure 2.10).

Table 2.15 Classical assignment test results undertaken in GeneClass 2. ID = Individual sample ID; P SAF = probability individual assigns to South Africa; P ANG = probability individual assigns to Angola; SC = *S. cantharus*; SE = *S. emarginatum*; B= Benguela (Angola); N= Namibe; FB = False Bay; MB = Mossel Bay; * indicates individuals with less than three loci genotyped (i.e. missing loci).

	P SAF	P ANG	ID	P SAF	P ANG	ID	P SAF	P ANG
SCB1*	0.1477	0.8823	**SCN9**	0.0135	0.6105	**SEFB15**	0.8922	0.0234
SCB2	0.0009	0.8594	**SCN10**	0.0334	0.8265	**SEFB16**	0.8944	0.0061
SCB3	0.0208	0.6346	**SCN11**	0.0917	0.9289	**SEFB17**	0.9921	0.0169
SCB4	0.0099	0.8504	**SCN12**	0.1121	0.5022	**SEFB18***	0.4844	0.0704
SCB5	0.0438	0.9893	**SCN13**	0.0008	0.6340	**SEFB19**	0.9815	0.0217
SCB6	0.0022	0.8541	**SCN14**	0.1593	0.9928	**SEFB20**	0.2963	0.0343
SCB7	0.4385	0.5712	**SCN15**	0.3766	0.6348	**SEFB21**	0.1641	0.0000
SCB8	0.1854	0.9326	**SCN16**	0.0573	0.9907	**SEFB22**	0.6488	0.0066
SCB9	0.0092	0.8680	**SCN17**	0.0393	0.1860	**SEFB23**	0.9474	0.0051
SCB10	0.1804	0.5779	**SCN18***	0.8472	0.8093	**SEFB24**	0.7289	0.0000
SCB11	0.0094	0.8220	**SCN19**	0.3102	0.8538	**SEFB25**	0.9814	0.0047
SCB12	0.0021	0.8610	**SCN20***	0.7193	0.9621	**SEFB26**	0.9858	0.0004
SCB13	0.2380	0.6950	**SCN21**	0.0000	0.0817	**SEFB27**	0.9592	0.1521
SCB14	0.0020	0.7640	**SCN23**	0.1353	0.8560	**SEMB1**	0.6801	0.0000
SCB15	0.2726	0.9390	**SCN24**	0.0075	0.9686	**SEMB2***	0.3450	0.0000
SCB16	0.4929	0.8024	**SCN25**	0.0010	0.9695	**SEMB3***	0.9365	0.0004
SCB17	0.0146	0.8400	**SCN26**	0.0227	0.7780	**SEMB4**	0.7919	0.0000
SCB19	0.0118	0.6761	**SCN27***	0.0090	0.6728	**SEMB5**	0.6811	0.0275
SCB20	0.0014	0.9368	**SCN28**	0.0175	0.8853	**SEMB6***	0.5395	0.0079
SCB21	0.0002	0.1224	**SCN29**	0.0273	0.8510	**SEMB7***	0.7139	0.0729
SCB23	0.0058	0.6875	**SCN30**	0.0779	0.5709	**SEMB8**	0.9826	0.0464
SCB24*	0.9802	0.9317	**SEFB1**	0.8979	0.0019	**SEMB9**	0.7035	0.0021
SCB25*	0.2993	0.6421	**SEFB2**	0.4883	0.0000	**SEMB10**	0.9376	0.0272
SCB26	0.0002	0.7645	**SEFB3***	0.4610	0.0020	**SEMB11**	0.8667	0.0029
SCB27	0.1977	0.7258	**SEFB4**	0.8620	0.0000	**SEMB12**	0.5347	0.2555
SCB28	0.0563	0.9547	**SEFB5**	0.8885	0.0000	**SEMB13**	0.8058	0.0231
SCB29	0.0097	0.8283	**SEFB6**	0.8581	0.0407	**SEMB14**	0.4618	0.2309
SCN1	0.0157	0.9535	**SEFB7**	0.9893	0.0336	**SEMB15**	0.9857	0.0333
SCN2	0.0008	0.6195	**SEFB8**	0.7513	0.0277	**SEMB16**	0.9526	0.0141
SCN3	0.0192	0.9829	**SEFB9**	0.8087	0.0236	**SEMB17**	0.4801	0.0244
SCN4	0.0192	0.5829	**SEFB10**	0.1660	0.0000	**SEMB18**	0.8931	0.0006
SCN5	0.0073	0.8879	**SEFB11**	0.9791	0.0140	**SEMB19**	0.9786	0.0000
SCN6	0.0674	0.4416	**SEFB12**	0.5053	0.0002	**SEMB20**	0.6857	0.0000
SCN7	0.0835	0.9967	**SEFB13**	0.8743	0.0179	**SEMB21**	0.9797	0.0033
SCN8	0.0124	0.9215	**SEFB14**	0.9724	0.1273	-	-	-

Table 2.16 Pairwise F_{ST} values between *Spondyliosoma* spp. samples based on four microsatellite loci calculated in FSTAT, significance assessed after 10000 permutations and Bonferroni correction for multiple comparisons. Sample ID as per Table 2.1. Significant values are in bold.

	BEN	NAM	FAL
NAM	-0.009	-	
FAL	**0.103**	**0.098**	-
MOB	**0.105**	**0.094**	-0.007

Table 2.17 Pairwise F_{ST} values between *Spondyliosoma* spp. samples based on four microsatellite loci calculated in FreeNA: Below the diagonal- uncorrected F_{ST}; and above the diagonal- F_{ST} after correction for null alleles. Sample ID as per Table 2.1. Note significance by bootstrapping could not be performed on only four loci.

	BEN	NAM	FAL	MOB
BEN	-	-0.006	0.094	0.087
NAM	-0.009	-	0.103	0.093
FAL	0.109	0.104	-	-0.003
MOB	0.106	0.096	-0.006	-

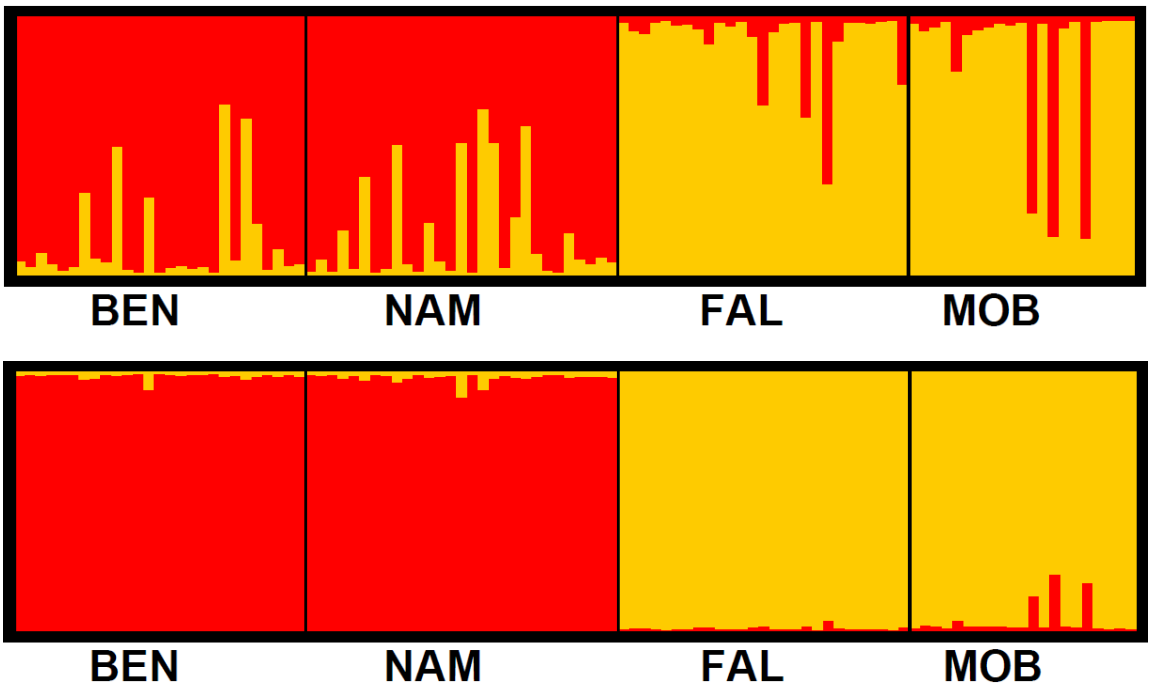

Figure 2.9 Number of genetic clusters observed within *Spondyliosoma spp* populations across the Benguela region. Assignment values for each individual fish (represented by bars) obtained from STRUCTURE, based on genotypes from four nuclear microsatellite loci, for K = 2. **Top** is the plot not assuming priors and **Bottom** is the plot assuming priors. Red= Angolan cluster, Yellow= South African cluster. Sample ID as per Table 2.1.

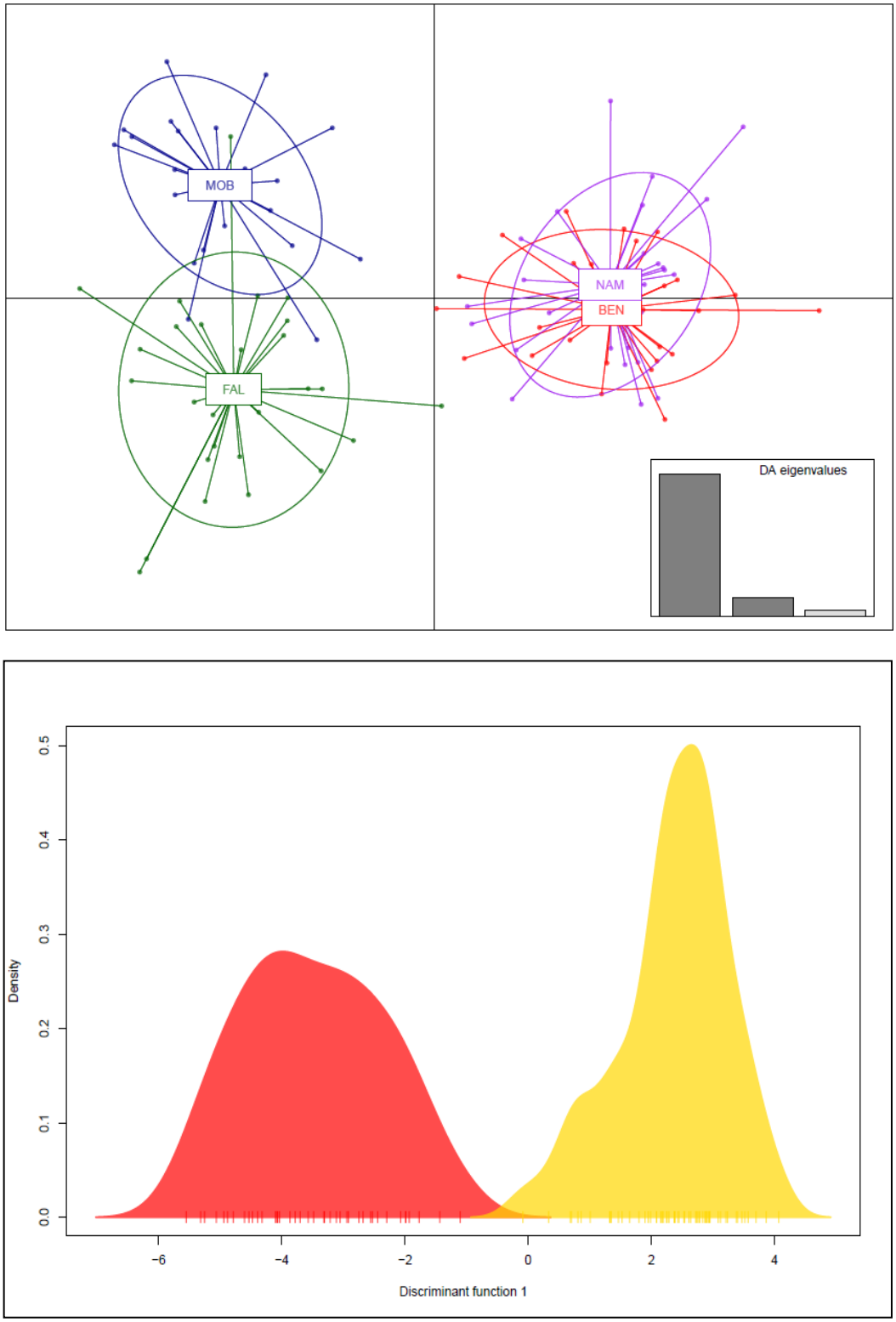

Figure 2.10 TOP: scatter plot of individuals on the two principal components of the DAPC with groups defined *a priori* as per sample site. The graph represents individuals as dots and the groups as inertia ellipses. Eigenvalues of the analysis are displayed in inset. Sample codes are as per Table 2.1. BOTTOM: Graph showing results of the DAPC obtained with the *find.clusters* option where clusters= 2, with a single discriminant function displayed on the x- axis, (with individuals represented as dashes) whilst density of individuals is plotted on the y -axis. Red refers to *S. cantharus* Angolan sampling sites (Benguela and Namibe). and Yellow refers to South African sampling sites of *S. emarginatum* (False Bay and Mossel Bay).

2.3.6 Power Analysis

POWSIM analysis indicates that the microsatellite data (average sample size = 26) has a low Type I error (Fisher P = 0.043) and a high probability (Fisher P= 0.984) for detecting differentiation at F_{ST}= 0.010. Both mtDNA markers also conferred a low Type I error probability, (Fisher P: CO1= 0.048; CR= 0.049). The mtDNA markers had a lower power to detect differentiation at F_{ST}= 0.010 (Fisher P: COI= 0.219; CR= 0.175) than the microsatellite data set; for both mtDNA markers power only reached the 95% threshold when F_{ST}= 0.07 (Fisher P: COI= 0.977; CR= 0.969).

2.4 Discussion

The most prominent feature of the genetic (principally mtDNA) data is the pronounced reciprocally monophyletic divergence of four clades within *Spondyliosoma* corresponding geographically to contemporary Mediterranean (represented by Turkey), North East Atlantic, Angolan and South African population samples. Estimates of genetic differentiation among all four population groups are very high (mtDNA Φ_{ST} >0.92). The division between Angolan and South African *Spondyliosoma* is also supported by the nuclear microsatellite data (F_{ST}> 0.09). This pattern of differentiation and divergence supports not only the existing species distinction of the endemic South African *S. emarginatum* but also indicates a possible cryptic species complex within *S. cantharus*.

2.4.1 Genetic divergence of Mediterranean / Atlantic *Spondyliosoma*

Reconstructed network-based and tree-based phylogenetic relationships, and mean levels of sequence divergence, amongst the four identified clades suggest greater (i.e. most likely earlier) divergence between the Atlantic and Mediterranean *Spondyliosoma* than among the Atlantic clades. Mean sequence divergence between Mediterranean and Atlantic *Spondyliosoma* is the highest at 2.7-3.1% whereas amongst the three Atlantic clades it is 2.0-2.3%, compared to 0.04-0.19% within the Atlantic clades and 0.34% within the Mediterranean clade. Molecular clock methods suggest divergence of the Mediterranean clade from the three Atlantic clades around 2.25-2.58 Ma. A Mediterranean-Northeast Atlantic split has been previously identified in *Spondyliosoma* by Bargelloni *et al.* (2003), who estimated the divergence to have occurred approximately 1.5 Ma (based upon CR sequence data). Bargelloni *et al.* (2003) also identified a similar major phylogeographic Atlantic-Mediterranean break in another

coastal Sparid, *Dentex dentex*, with an estimated divergence of 2 Ma. Although these levels of divergence are high for intra-specific comparisons, Bargelloni *et al.* (2003) did not suggest that they were associated with cryptic speciation but that they were typical of the major phylogeographic breaks between Atlantic and Mediterranean populations of some species.

Suggested divergence time between Atlantic and Mediterranean *Spondyliosoma* correlates with global cooling around the Miocene-Pleistocene transition and associated known environmental destabilisation(s) occurring within the Mediterranean Sea over the last 2 Myrs, with habitat fragmentation in the western Mediterranean and Alboran Sea, along with recurrent short periods of the Atlantic and Mediterranean being physically isolated during glacial maxima (Patarnello *et al.*, 2007). These conditions likely created large stretches of inhospitable habitat for *Spondyliosoma* in the North East Atlantic near the species northern range edge (Portugal-UK) and in the western Mediterranean, which are likely to have initiated the isolation of the two clades. The recognised present day oceanographic barrier to dispersal and migration of fish between the Atlantic Ocean and Mediterranean Sea is the Almeria-Oran frontal system (Perez-Losada *et al.*, 2007), which is a likely mechanism for maintenance of the observed divergence between *Spondyliosoma* populations in the region.

2.4.2 Genetic divergence among Atlantic populations of *Spondyliosoma*

Atlantic *Spondyliosoma* are resolved into three monophyletic clades corresponding geographically to North East Atlantic, Angolan and South African population samples. Phylogenetic reconstructions failed to resolve a clear dichotomy amongst these clades, instead signifying a polytomy (Figures 2.3 to 2.8), suggesting initial separation of the three clades at similar times around 1.67- 1.92 Ma (COI data). The identified groupings are supported by the present observed distribution of *Spondyliosoma* (see Figure 2.1), where there are clear breaks in the distribution between the North East Atlantic and Angolan *Spondyliosoma* and a break in distribution between Angola and South Africa around the position of the Benguela Current upwelling area. Whilst the accuracy of species distribution maps can be patchy and requires further data (Franklin, 2010), the model of genetic divergence observed in this present study supports the species distribution map currently shown by the Global Biodiversity Information Facility (GBIF; Figure 2.1) as being accurate, with a break in *Spondyliosoma* distribution around the equator between West Africa and Angola. Recent expeditions to Namibia

failed to find *Spondyliosoma* in the cold Benguela Current waters (W. Potts, pers comm., 2014), again supporting the GBIF distribution map.

Whilst the Mediterranean / Atlantic divide and the role of the Benguela Current (see below) as biogeographical barriers are established and relatively well studied, there is little research suggesting a barrier to dispersal in coastal fishes between Angola and the North East Atlantic. Likely extrinsic barriers to movement of *Spondyliosoma* between Angola and the North East Atlantic are the South Atlantic Equatorial Current, Guinea Current and Angolan Current. All three currents are warm equatorial currents, with the South Atlantic Equatorial Current deflecting from the West African coast across the Atlantic to the coast of South America (Peterson and Stramma, 1991). The Guinea Current travels southwest along the Gulf of Guinea coast and also deflects westward into the central Atlantic, as does the Angolan Current from north-central Angola (Peterson and Stramma, 1991). These equatorial currents may pose as a barrier to dispersal in *Spondyliosoma* either physiologically by being too warm for *Spondyliosoma* adults and / or larvae to survive, or being an ecological barrier due to the differing species assemblages associated with the warm currents leading to a lack of suitable food items and / or being outcompeted (Sherman and Hempel, 2008). The deflection of these currents offshore into the central Atlantic also may result in the deposition of any *Spondyliosoma* larvae away from suitable coastal habitat into unsuitable Open Ocean. It should be noted however that larval *Spondyliosoma* tend to reside within the nest environment and actively seek the inshore habitat once in excess of 7 cm in length (Pawson, 1995), suggesting any such offshore displacement action unlikely. Whichever factors are responsible for preventing gene flow; such gene flow restrictions have been present and persistent for the past 1.5-2.0 Myrs. Antitropical distributions are observed in many Pacific fishes (Burridge, 2002), and in Atlantic Anchovies (*Engraulis*; Grant and Bowen, 2005) suggesting a prominent role for the tropical barrier in many marine fishes.

2.4.3 The Benguela Current as a biogeographic boundary

Populations of *Spondyliosoma* in Angola and South Africa are clearly highly diverged from one another, both in terms of present breakdown of gene flow (significant nuclear microsatellite allele frequency differentiation) and long-term isolation and genetic divergence (mtDNA clade divergence). Both mitochondrial markers identify substantial mean sequence divergence between Angolan and South African populations (2.3 % for

COI, 7.7 % for CR), which equates to divergence from a common ancestral population approximately 1.92 Ma (COI) to 2.13 Ma (CR). Nuclear variation (analysed by F_{ST}, and Bayesian clustering/assignment methods) revealed significant differentiation (F_{ST}> 0.09) between Angolan and South African samples consistent with absence of present (or recent historical) gene flow, and with the reciprocal monophyly observed in the mtDNA. Microsatellite F_{ST} values have also likely been downwardly biased, compared to mtDNA Φ_{ST} values, by their high levels of polymorphism (Hedrick, 1999) and potentially by homoplasy (Estoup and Angers, 2000; Estoup *et al*., 2002). The Bayesian assignment run in STRUCTURE using the no *locprior* model indicated that some South African individuals clustered with the Angolan genetic cluster, potentially indicating migration, whilst using the locprior model all individuals were assigned to the genetic cluster representing their geographical region of origin (Figure 2.9). However, migration of a few individuals from Angola to South Africa would seem unlikely: firstly, the locprior model has been shown to give higher resolution in such analyses without biasing results (Hubisz *et al*., 2009); and secondly, alternative assignment tests based on genetic distances reported assignment of all individuals to their region of sampling. Such assignments methods do not rely on HWE, suggesting that the poorly clustered individuals in the STRUCTURE no locprior model may be due to null allele effects (Carlsson, 2008). Additionally the DAPC results (which again are not reliant on HWE) identify two clear clusters corresponding geographically to Angola and South Africa with little to no overlap between the regions (Figure 2.10) suggesting there is no migration between the two regions. Collectively the data indicate little if any recurrent dispersal/gene flow between the two regions, likely since the time of initial divergence approximately 2 Ma.

The estimated timing of divergence between Angolan and South African *Spondyliosoma* is in agreement with previous studies of population and species isolation and divergence for inshore fish species across the Benguela Current region: Henriques (2012) identified a divergence time of ~2 Ma between two sister species of Croaker *Argyrosomus japonicus* (South Africa) and *Argyrosomus coronus* (Angola); similarly Henriques *et al.* (2014) identified a cryptic speciation event between Geelbek *Atractoscion aequidens*, populations in Angola and South Africa, again with a divergence date of approximately 2 Ma. Whilst the Benguela Current presents a formidable barrier for some species, it has been at least historically permeable for some.

Both *Diplodus capensis* and *Lichia amia* have more recent times of population divergence across the Benguela region (367 Kya (Henriques, 2012) and 222 Kya (Henriques *et al*., 2012) respectively). The long term isolation of Angolan and South African *Spondyliosoma* suggests that this fish may be more sensitive to the features of the Benguela Current which restrict dispersal than some other fishes.

The perennial upwelling cells within the Benguela Current system (and others worldwide) are often cited as being potential barriers to dispersal, particularly for larval stages of coastal fishes or other taxa, whereby larvae drifting into the upwelling area are transported offshore into the (inhospitable) open ocean and thus do not recruit successfully into the next generation (Waters and Roy 2004; Lett *et. al.*, 2007). The life history characteristics of *Spondyliosoma*, however would suggest that potential for larval dispersal is very limited and that dispersal will be effected more by adults; as such the offshore currents generated by upwelling cells (particularly the perennial Lüderitz cell) would seem unlikely to present a major physical dispersal barrier to this (relatively) large marine fish. A more likely barrier to adult (and juvenile) dispersal across the upwelling cells is the associated low sea surface temperatures, which might represent a physiological barrier to adults, discouraging them from moving into the upwelling region. Likewise the Benguela Current region may be an inhospitable environment for reproduction to occur as water temperatures may be too cold for successful gamete development or survival of embryonic/larval stages. The continental shelf along the Namibian coast is also narrow compared to that found in South Africa and Angola, limiting available habitat space and suitable nesting sites, (Siesser and Dingle, 1981). Anecdotal evidence from local fishermen in Aberystwyth (Wales UK) suggests that *S. cantharus* appear to use the same nesting site areas year after year; this could imply a certain degree of breeding philopatry in *Spondyliosoma* generally. So even if adult *Spondyliosoma* traverse the Benguela Current upwelling area they may not spawn in the region, instead migrating back to their natal region. Finally, the Benguela Current is a highly productive region, so it is possible that *Spondyliosoma* are outcompeted in the region by other fishes, or that specific food and other aspects of the niche of *Spondyliosoma* are missing in the region. On balance the most obvious feature likely to discourage movement and dispersal between the northern and southern Benguela sub-systems by *Spondyliosoma* adults and juveniles, as suggested for other warm temperate fish species across SW Africa coasts (Henriques, 2012), is the presence

of the strong offshore currents and cold sea surface temperatures of the Lüderitz perennial upwelling cell, an oceanographic feature that became established at the same time as the population isolation indicated by the present study data (~2 Ma: Shannon, 1985; Diester-Haas *et. al.* 1988).

Tests for population structuring within regions (Angola and South Africa) for *Spondyliosoma* using the mitochondrial genetic markers identify a pattern of panmixia within regions: Φ_{ST} values within regions are low and non-significant whilst comparisons between the two regions were exceptionally high for a marine fish and highly significant (Tables 2.9 and 2.10), with an overall Φ_{ST} for Angolan / South African comparisons of 0.97 (COI) and 0.95 (CR). Despite the small number of microsatellite loci included in this study power analysis identified that the data set could successfully identify F_{ST} values as low as 0.01 successfully 98.4% of the time. Nuclear microsatellite loci are concordant with the mtDNA markers, with apparent panmixia within regions where F_{ST} values are effectively 0 (Tables 2.16 and 2.17). The present study is missing sample(s) from the KwaZulu-Natal region in South Africa and as such some level of genetic population structuring for *S. emarginatum*'s extant range cannot be ruled out. Adaptive genetic markers may also reveal finer levels of structuring of interest to fisheries management (Cadrin *et al*., 2013; Milano *et al*., 2014) in both regions, but such analyses are beyond the remit of this study. These limitations notwithstanding it is clear that within regions there is limited genetic structuring indicating that dispersal potential is realised within regions but not across the Benguela Current barrier. Dispersal most likely occurs through adult movements as *Spondyliosoma* are demersal spawners with larvae staying within the nest environment until they are approximately 7.8cm in length after which they disperse locally into the inshore area until they join the adult stock at 2 years old (Pawson, 1995). Dispersal is documented in adult *S. cantharus* which are known to migrate seasonally within the English Channel (Pawson, 1995) and Cardigan Bay. However in southern Africa dispersal potential appears to be only realised within regions and not between the northern and southern Benguela Current regions. The Benguela Current thus appears to act as a barrier to dispersal both in the present day (no shared haplotypes and a high degree of microsatellite structuring) and historically (to accumulate the large genetic distances between the two populations).

2.4.4 Demographic history

All three Atlantic clades show signals of historical population expansions, whilst in keeping with the study by Patarnello *et al*. (2007) the Mediterranean clade does not. In the Atlantic clades COI estimates of time since expansion are much older (106-227 Kya) than those inferred from the CR (26-42 Kya). Such a discrepancy could be possible for two reasons. Firstly the COI exhibits a limited amount of diversity compared to the CR, making it inadequately variable and unsuitable for studying demographic history (Harpending *et al*., 1998; Grant, 2015), which may account for the large 95% CI associated with the COI (Table 2.3). Secondly, the COI sequences should show lower diversity since it accumulates mutations at a much slower rate than CR; therefore it is plausible that the COI has identified older (pre LGM) demographic expansions which are lost in the more highly variable CR sequences, however pre LGM population expansions are considered controversial given the numerous sources of error in mismatch analyses and necessitating in this scenario that the LGM did not affect the COI marker (see Grant, 2015 for review)

Estimates of time since expansion from the CR data set identified expansions dating from the last glacial period. The oldest estimated time since expansion was identified in South African clade (Texp = 42 Kya), then the North East Atlantic clade (Texp = 36 Kya) and the Angolan clade exhibiting the most recent date since expansion (Texp = 26 Kya). Expansions dating from the last glacial period are found in many coastal fishes in southern Africa, such as *Merluccius paradoxus* (Texp = 23 Kya; von der Heyden *et al.*, 2007),*Merluccius capensis* (Texp = 6 Kya; von der Heyden *et al*., 2010) and *Caffrogobius caffer* (Texp = 45-138 Kya; Neethling *et al.*, 2008). Such population crashes and subsequent expansions are most likely due to the unfavourable conditions endured in the region during glacial periods with cooler sea surface temperatures, lower sea level and reduced marine productivity (Marlow *et al*., 2000; Jahn *et al*., 2003).

The Angolan *Spondyliosoma* clade shows the most extreme bottleneck/ expansion signal with its clear star-shaped haplotype network (Figures 2.3 and 2.5), typically found following an extreme population contraction and expansion from a surviving common haplotype, and displaying the most recent time since expansion. The Angola region has a much narrower continental shelf compared to South Africa (with the large extensive Agulhas Bank) limiting available habitat for *Spondyliosoma* particularly during sea level falls during glacial cycles (Ramsay and Cooper, 2002). The range of

Angolan *Spondyliosoma* is also restricted, being bound by the warm equatorial currents to the north and the cold Benguela Current to the south. Both of these currents are known to have been in place for the last 2 Myrs (i.e. since the last time of contact with the North East Atlantic and South African populations), suggesting that the *Spondyliosoma* clade in Angola has always been restricted in its range, therefore potentially magnifying any environmental perturbation effects on population size compared to the wider ranges inhabited by the other populations.

2.4.5 Species question

The mtDNA reciprocal monophyly and nuclear divergence between the Angolan and South African samples support recognition of the South African endemic *S. emarginatum* as a distinct species. However, comparable reciprocal monophyly and sequence divergence in the mtDNA markers was reported for the other two Atlantic clades (Angola and North East Atlantic), as well as an even higher level divergence of the Mediterranean Clade. Ideally a speciation study requires several loci (i.e. mtDNA / nuclear loci, preferably a combination of both neutral and non-neutral markers) and incorporating species morphology, life history and geographical distribution, with congruence across these methodologies providing strong support for species delimitation (Knowles and Carstens, 2007; Schlick-Steiner *et al*., 2010; Carstens *et al*., 2013). Nuclear genetic and morphological data in the present study are lacking (largely since this was an unexpected result) for the three proposed cryptic species in *S. cantharus* but they are supported by their geographical distribution and mtDNA markers. In contrast the species delimitation identifying the South African endemic *S. emarginatum* is additionally supported by four nuclear microsatellite loci as well as geographical isolation. Furthermore initial results of a study into morphological and ecological differences between South African and Angolan *Spondyliosoma* have found significant divergence between these two populations (J. Kruger, Rhodes University, pers. comm.). As such the *S. cantharus* / *S. emarginatum* species divide is highly robust and reinforces the notion that *S. emarginatum* be considered a species in its own right. This raises the question of whether the three diverged *S. cantharus* clades also warrant species status?

Under the phylogenetic species concept (PSC) the three reciprocally monophyletic *Spondyliosoma* clades satisfy the criteria for speciation (Cracraft, 1983; Hudson and Coyne, 2002). Interclade divergence was also estimated to be far greater than intraclade

divergence (Tables 2.7 and 2.8): interclade sequence divergence estimates should always be greater than estimated intraclade divergences in 'true' species (Held, 2003). Furthermore the ratios of interclade / intraclade divergence are in line with a proposed 10 X rule, where between-species sequence divergence is at least an order of magnitude greater than within-species (Hebert *et. al.*, 2004), a rule which is commonly applied to vertebrate taxa and often applied to fishes (e.g. Khedkar *et. al.*, 2014). Applied to the present study the mean intraclade sequence divergence for COI in the four *Spondyliosoma* clades is 0.18%, signifying that all four clades would warrant species status under the '10 X rule' since the minimum observed divergence between the clades is 2% (Angola-NE Atlantic, Table 2.7). Likewise, all three clades identified by the CR sequence data would warrant species status under the '10 X' rule with a cut off of 4.6 % (Table 2.8). Previous research into levels of genetic divergence observed in COI associated with delineating marine coastal fish species identified an average of 0.39% within-species K2P distance, compared to an average 8.91% mean K2P distance between congeneric species (Landi *et. al.*, 2014). The minimum interspecific genetic distance within a genus observed by the Landi *et. al.* (2014) study was 1.09% in the genus *Trachurus*. The observed levels of genetic distance in the *Spondyliosoma* clades can thus be interpreted as being much lower than the average level of divergence between congeneric species, but not the lowest observed for recognised marine teleost species.

Whilst the PSC and threshold criteria can provide useful estimates of species delimitation there are major shortcomings. For example a monophyletic criterion used singly could identify several further taxa within the clades already highlighted. Whilst exclusivity criteria have been heavily criticised for their arbitrary nature in what constitutes a species (Hudson and Turelli, 2003; Knowles and Carstens, 2007), both PSC and exclusivity criteria both also ignore the biological reality and complexity of the speciation process (often disregarding recently diverged taxa). The present study relies heavily on mtDNA, which acts as a single locus leaving the present study vulnerable to the stochasticity of the coalescent process that leads to speciation (Hudson and Turelli, 2003), and as such may not reflect the true species tree when further (nuclear) loci are included. However, use of neutral nuclear loci may not aid in identifying recently diverged taxa particularly given their large *Ne* compared to mtDNA markers (Zink and Barrowclough, 2008). This disparity may be even greater for Sparids, in a study by von

der Heyden and Connell (2012) neutral nuclear markers failed to resolve recognised species and in some cases species from different genera. As such any future genetic study into *S. cantharus* taxonomy will need to include non- neutral nuclear loci (perhaps utilising next generation sequencing technologies), alongside a life history and morphological study to fully resolve the taxonomic issues in *S. cantharus* raised by the present study.

In the absence of a taxonomic review, it is clear that the *S. cantharus* clades have been isolated for a considerable length of time and are on their own independent evolutionary trajectories. Whether or not they are distinct species, all three *S. cantharus* clades are of a monophyletic origin, with independent demographic histories and are geographically isolated, and therefore all three *S. cantharus* clades can be classed as Evolutionary Significant Units (Moritz, 1994; Crandall *et al.,* 2000). Therefore it would be prudent that all three lineages of *S. cantharus* should at least be considered as separate 'stocks' in terms of fisheries management and conservation.

2.4.6 Relevance to conservation and fisheries

This study identifies four clear lineages or 'stocks' in *Spondyliosoma*, which appear to have been separated from each other for at least 1.5 Myr. As a fishing resource (and at a rather coarse level) *Spondyliosoma* should be considered as four separate stocks corresponding to the North East Atlantic, Mediterranean, Angola and South Africa, with little or no migration between them. The hermaphroditic nature and breeding behaviour of *Spondyliosoma* spp. could make it more susceptible to overfishing, with the largest individuals within a population typically being male and acting as nest guards protecting the eggs and early larval stages from predation (Pawson, 1995). Also of note in this study is the apparent relatively small population size of *Spondyliosoma* in Angola where they appear to have undergone a comparatively more severe bottleneck than in the other three regions (possibly this population has been constrained in growth for hundreds of thousands of years). This restriction could make the Angolan *Spondyliosoma* population particularly sensitive to loss of genetic biodiversity through overfishing.

The Benguela Nino is a periodical phenomenon equivalent to the Pacific El Nino, during which warm equatorial currents push straight through Angolan waters into the northern Benguela Current. As this study has highlighted that both the equatorial and

Benguela current have acted as barriers to *Spondyliosoma* dispersal over at least the last 1.5 Myr this could have profound effects on breeding success of *Spondyliosoma* during Nino years. Current climate change models and the link to Pacific El Nino suggest the Benguela Nino could continue with increased regularity and severity (Richter *et. al.,* 2010) along with increased strength and cooling of the Benguela Current upwelling system and warmer equatorial currents; it is possible that the relatively small suitable habitat for Angolan *Spondyliosoma* could be further restricted geographically and environmentally over the next 100 years.

2.5 Conclusions

From a phylogenetic perspective there is no support for the present description for just two species in *Spondyliosoma*, with the present study identifying four equally divergent phylogenetic clades corresponding geographically to the Mediterranean, North East Atlantic, Angolan and South African parts of the range. Whilst such a pattern is supportive of the species status of the South African endemic *S. emarginatum*, it is also suggestive of cryptic speciation within *S. cantharus*. Phylogenetic analyses identified the earliest break between the four clades was between the three Atlantic clades and the Mediterranean clade, dating to approximately 2.25-2.58 Ma. This was then subsequently followed by a polytomic split (i.e. occurring concurrently) between the three Atlantic clades approximately 1.67- 1.92 Ma. Whilst this study does not have sufficient evidence for definite species status of these four clades it is clear that all four clades conform to a definition as ESUs, are on their own independent evolutionary trajectories, and perhaps should be considered as independent genetic taxa (i.e. species). It is suggested that a comparative study of morphology, life history and ecology of the four clades should be undertaken to establish a new systematic description of the genus. A study of the historical demography of the four clades indicated population expansions in the three Atlantic clades but no support for such an expansion in the Mediterranean clade. Within the Atlantic the Angolan clade exhibited the most recent and pronounced signal of population contraction and expansion, which is most likely due to historical bottlenecks from isolation and limited habitat available to *Spondyliosoma* in Angolan waters. Benguela Nino effects could have serious implications for *Spondyliosoma* in Angola, affecting fishery size and sustainability.

Chapter 3: Comparative phylogeography of two Sparid fishes: *Lithognathus mormyrus* and *Sarpa salpa.*

3.1 Introduction

Comparative phylogeography is defined as the study of the effects of evolutionary history and biogeography on the distribution of genetic diversity in co-distributed species (Avise 2009; Gutiérrez-García and Vázquez-Domínguez, 2011; Avise, 2016). By comparing phylogeographic structure among co-distributed taxa with similar ecological preferences and dispersal abilities, the relative roles of ancient and more recent extrinsic (e.g. glaciation) and intrinsic (e.g. behaviour) factors can be partitioned. Recent comparative phylogeographic studies have been performed in a number of marine areas, including the North Eastern Pacific (Kelly and Palumbi 2010; Marko *et al*., 2010), Atlantic-Mediterranean (Patarnello *et al*., 2007), the North Atlantic (Maggs *et al*., 2008) and Southern Australia (Ayre *et al*., 2009). These have provided considerable insight into the interplay of historical biogeography, oceanography and ecological factors in shaping marine biodiversity.

The previous chapter investigated the role of putative biogeographic barriers to dispersal along the west coast of Africa in *Spondyliosoma*, finding them to be apparent absolute barriers promoting cryptic speciation. However other marine coastal fishes are distributed throughout the region occurring either side of the Benguela Current barrier with no suggestion of speciation or evolutionary divergence. This suggests that the Benguela Current barrier may be permeable at varying levels for some coastal fishes, allowing dispersal and gene flow both historically and in the present day (Henriques 2012; Henriques *et al*., 2012). Two such species occurring along the west coast of Africa are *Lithognathus mormyrus*, the striped Seabream (Figure 3.1) and *Sarpa salpa*, the Salema Porgy (Figure 3.2). Both are coastal Sparids with similar life history traits and geographical ranges. Both species have warm temperate coastal distributions ranging from South Africa (or the Indian sub-continent in the case of *L. mormyrus*) to

the Mediterranean (Figures 3.1 and 3.2). There have been no suggestions from taxonomists of any major taxonomic structuring within either of the species, both being recognised as *L. mormyrus* and *S. salpa* throughout their ranges. Based on their ecological similarity and the possibility for cryptic diversity both species are prime candidates for a comparative phylogeographic study (Guitierrez Garcia and Vazquez-Dominguez, 2011).

3.1.1 *Lithognathus mormyrus*

Lithognathus mormyrus occurs throughout the Mediterranean Sea and the North East Atlantic Ocean (as far north as the Bay of Biscay). Along the west coast of Africa it occurs sporadically to Angola and is absent from the cold Benguela Current waters from central Namibia to north-western South Africa, where its distribution continues throughout South Africa, Mozambique and the western Indian Ocean (Figure 3.1). *L. mormyrus* adults prefer sandy bottom substrates and sea grass beds, occurring to depths of 150m and are known to occasionally enter brackish waters (Bauchot and Hureau, 1986). *L. mormyrus* tend to be found together forming considerable schools (Heemstra and Heemstra, 2004). It is carnivorous, feeding on worms, molluscs and small crustacea. There is some observed partitioning of habitat and feeding behaviour of juvenile and adult *L. mormyrus.* Juveniles tend to occur in shallow waters (<10 m), nurseries are found in large bays in the Eastern and Western Cape of South Africa (Lasiak, 1981 and 1986), whilst adults tend to occur in deeper waters (>10 m). Juvenile *L. mormyrus* tend to feed upon detritus (Gushchin *et al.*, 2013), and older individuals tend to be more generalist.

Lithognathus mormyrus are generally considered to be protandrous hermaphrodites; however some studies have found evidence that they are monandric protogynous hermaphrodites, meaning protogynous hermaphrodites with only juvenile females, where a certain percentage of females change into (terminal) males. *Lithognathus mormyrus* typically have a maximum lifespan of 12 years (Kraljevic *et al.*, 1996) attaining a maximum length of 55 cm (TL; Bauchot, 1987), but are typically about 30 cm (TL; Bauchot and Hureau 1990).

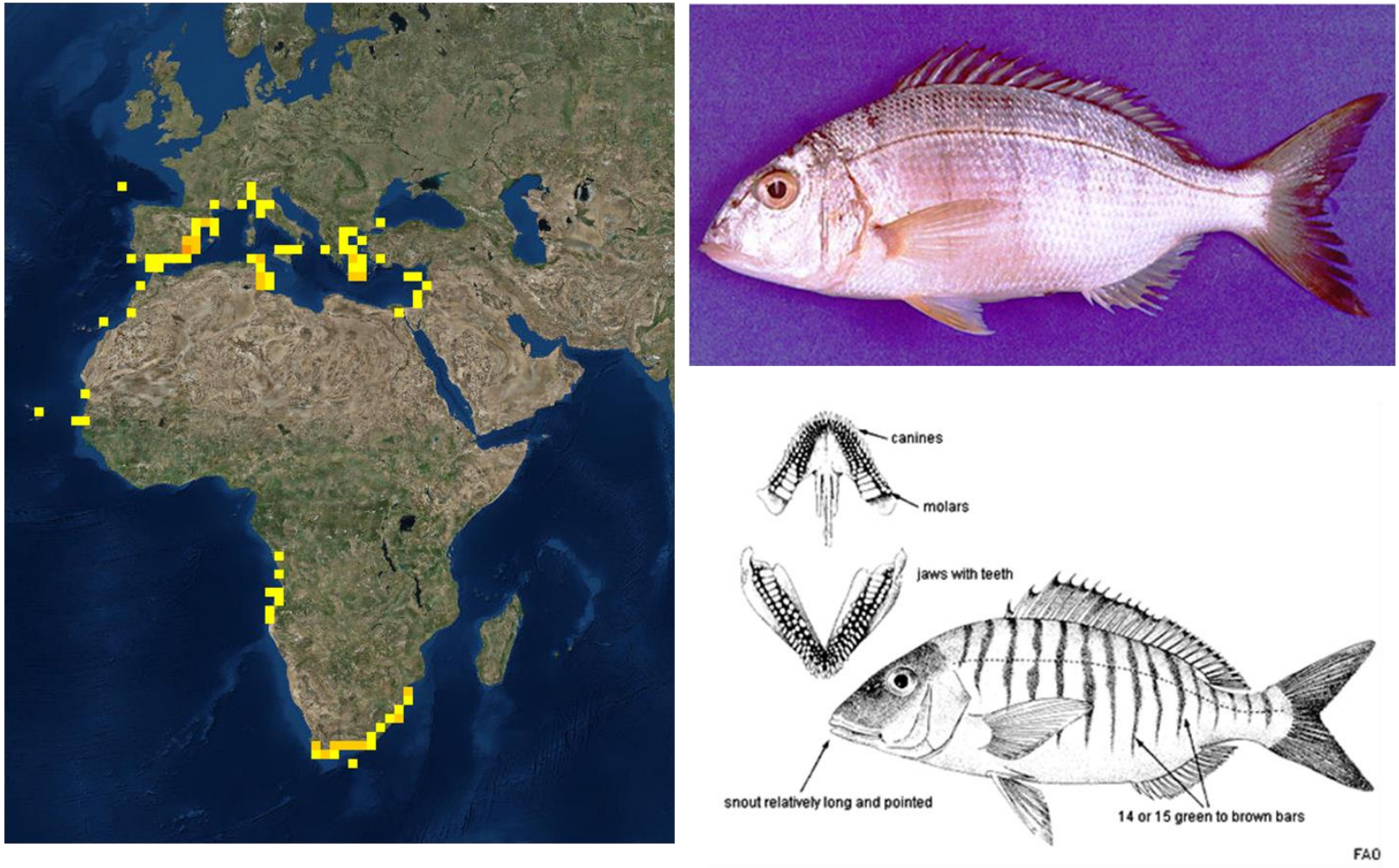

Figure 3.1 *L. mormyrus* distribution map, Red / Orange indicate a higher number of point localities recorded. Specimen and diagrammatic representation of *L. mormyrus* (left). Map taken from GBIF: http://www.gbif.org/

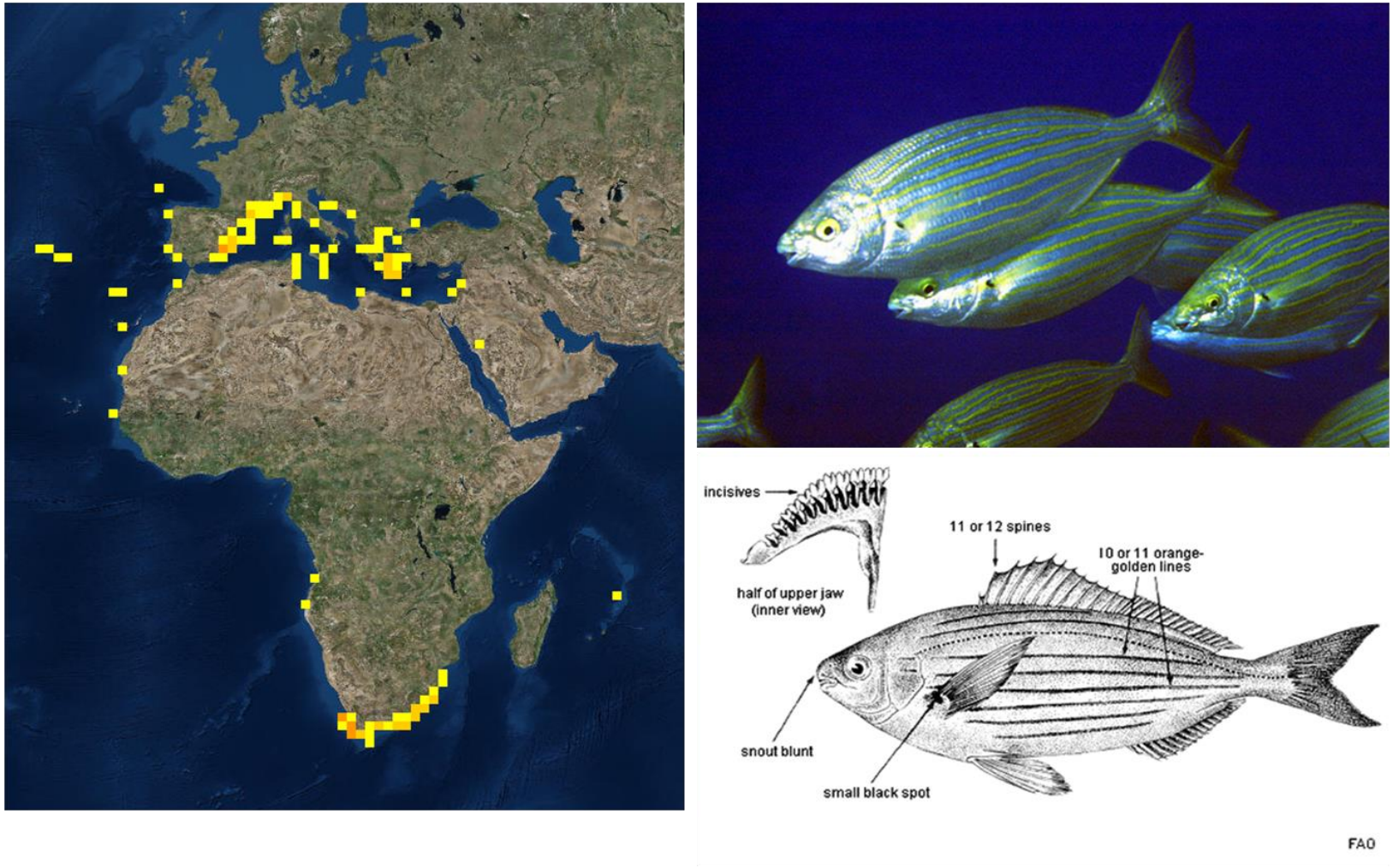

Figure 3.2 *S. salpa* distribution map, Red / Orange indicate a higher number of point localities recorded. Specimen and diagrammatic representation of *S. salpa* (left). Map taken from GBIF: http://www.gbif.org/

3.1.2 *Sarpa salpa*

Sarpa salpa is distributed throughout the Mediterranean, North East Atlantic (as far north as the Bay of Biscay), the west coast of Africa, South Africa and into the south western Indian Ocean (Mozambique; Figure 3.2), with the same break in distribution as *L. mormyrus* through the cold waters of the Benguela Current. It is herbivorous, grazing on algae, inhabiting areas that have rock or sandy bottoms covered in seaweed. *Sarpa salpa* is known to be an important grazer of sea grass meadows in the Mediterranean with a long gut to help digest the plant matter (Havelange *et al.*, 1997). Juveniles occupy tidal rock pools, sandy littoral zone, estuaries and shallow reefs in the Eastern and Western Cape in South Africa (Mann and Dunlop, 2013). Within South African waters juvenile *S. salpa* have nursery areas in the Western and Eastern Cape and when mature migrate to breed in the warmer waters of KwaZulu-Natal (Van der Walt and Mann, 1998). Although evidence is lacking of a return migration, some of the largest specimens of *S. salpa* caught within South Africa are from Cape Waters (Mann *et al.*, 2013). *S. salpa* has a maximum reported body size of 51 cm (SL; Bauchot, 1987) but with a common length of 30 cm (SL; Bauchot and Hureau, 1986). There is possibly some dichotomy of size in *S. salpa*, where both the maximum weight of 1.5kg (Canary Islands Villamil, 2002) and age (11 years, Villamil *et al.*, 2002) are recorded from the North East Atlantic region, whilst *S. salpa* in South Africa rarely exceed 30 cm, 0.7kg and with a maximum recorded age of 6 years (Mann and Dunlop, 2013).

Sarpa salpa is a protandric hermaphrodite (van der Walt and Mann, 1998, Villamil *et al*., 2002). Sex change normally occurs between 18 and 22 cm (FL), at an age of around 3 years. In Cape Verde two spawning periods have been identified, in spring (March to May) and in the autumn (September to November). In the Canary Islands spawning occurs from September to March, peaking in the midwinter (December to January; Villamil *et al.*, 2002). In KwaZulu-Natal (South Africa) spawning occurs in austral winter and spring (April to September), where spawning occurs over shallow sub tidal reefs (Joubert, 1981, Van der Walt and Mann, 1998, Connell, 2012). The age of 50% maturity in Kwazulu-Natal is at 1.5 years (Van der walt and Mann, 1998). The length at 50 % maturity for males is 14-15cm (FL) and 16-17cm (FL) for females (Joubert, 1981) and 14.5cm (FL) combined for both sexes (Van der Walt and Mann, 1998). *S. salpa* is reputed for its Ichthyoallyeinotoxic (hallucinogenic) properties (the Romans (allegedly) used *S. salpa* as a recreational drug), this can pose certain health risks to human

consumption (de Haro and Pommier, 2006; Bellasoued *et al*., 2012). The hallucinogenic properties are theorised to derive from toxic Dinoflagellates (living as epiphytes on the seagrass *Posidonia oceanica)* that are ingested and contaminate the flesh of the fish, accumulating in brain tissue (Helfrich and Banner, 1960; Chevaldonne, 1990; Bellasoued *et al.,* 2012).

3.1.3 Fisheries

Both species are harvested throughout their ranges. *Lithognathus mormyrus* is an important artisanal fisheries resource throughout its range and also supports a small scale commercial fishery. FAO catch statistics from 1996- 2006 ranged from 718-1135 tonnes in the Mediterranean with catches being relatively stable throughout the period. However in Angola *L. mormyrus* is a particularly important resource with some evidence of overfishing. The Angolan catch declined from 3353 tonnes in 2006 to 223 tonnes in 2007, the Angolan catch has recovered somewhat being 2030 tonnes in 2012. Globally the reported FAO catch of *L. mormyrus* in 2012 was 3632 tonnes. The IUCN red list assessed *L. mormyrus* as being of least concern; but recommended implementation of current and further fisheries management particularly where it is overfished, and that genetic research be undertaken to manage the species (Russell *et al*., 2014a).

Sarpa salpa is fished throughout its range by recreational anglers and subsistence shore fishers. It also supports a small irregular fishery. FAO landing statistics for the Mediterranean indicate a steady increase over the last 50 years peaking at 4000 tonnes in the early 1990s and have stabilised at around 2000 tonnes from 1996 onwards. Interestingly in the eastern Mediterranean there is circumstantial evidence that *S. salpa* are being outcompeted and displaced by the lessepsian migrants *Siganus luridus* and *Siganus rivulatus* both of which are herbivores feeding on sea grass (Lundberg *et al.,* 2004). The IUCN red list assessed *S. salpa* as being of least concern; however the report recommended fishing regulations due to the slow maturing reproductive strategy exhibited by *S. salpa*, as well as a genetic study to help manage the species (Russell *et al*., 2014b).

Previous genetic studies have focussed on the Mediterranean and North East Atlantic part of the *L. mormyrus* range. *L. mormyrus* was included by Alarcon and Alvarez (1999) in a study using Isozyme markers to identify Mediterranean Sparids. Arculeo *et al*. (2003) used several allozyme markers to assess genetic population structure from

several samples obtained from the Tyrrhenian Sea, Sicily, Adriatic Sea and the Aegean Sea, identifying panmixia within the region. Likewise Hammami *et al*. (2007) identified an absence of population structure in *L. mormyrus* across several sampling sites along the Tunisian coast using allozyme markers. Bargelloni *et al*. (2003) utilising mtDNA Control Region (CR) identified a major biogeographic split between Atlantic and Mediterranean *L. mormyrus*. Sala-Bozano *et al*. (2009) again utilising the mtDNA CR identified the same major split between the Mediterranean and Atlantic, whilst their additional samples detected no significant population structuring within the Mediterranean. Furthermore when utilising nuclear microsatellite data Sala-Bozano *et al*. (2009) found that whilst both populations are highly diverged (approximately 3.4 Ma), upon secondary contact individuals from each clade appeared to freely interbreed. There have been no previous published studies focussing on the population genetics and phylogeography of *S. salpa*.

3.1.4 Aims and Objectives

This study employed a range of genetic markers (mtDNA COI and CR, plus nuclear microsatellites) and analytical approaches (phylogeographic and population genetic) to reconstruct and compare the phylogeography of both *L. mormyrus* and *S. salpa* across their ranges. Data are specifically interpreted in the context of:

I. Exploring past and present genetic population structuring, divergence and gene flow between Angolan and South African populations of *L. mormyrus* and *S. salpa* (i.e. across the Benguela Current region);
II. Investigating how southern African demes relate genetically to North East Atlantic and Mediterranean populations.

3.2 Methods

3.2.1 Sampling and DNA extraction

Samples were collected from ten Angolan and South African locations (Tables 3.1 and 3.2; Figure 3.3). Additionally *L. mormyrus* samples were obtained from Portugal and Turkey to act as a European outgroup for analyses (Table 3.1), whilst only samples from Turkey could be obtained to act as an outgroup for *S. salpa* (Table 3.2). Samples were obtained from a mixture of recreational angling and local fish markets. A fin clip was removed from each individual and preserved in 95% ethanol. Total genomic DNA was extracted as per methods section in Chapter 2.

3.2.2 Mitochondrial DNA markers

Two mitochondrial DNA regions were amplified using PCR: Cytochrome Oxidase I (hereafter COI) and Control Region (hereafter CR). The COI region was amplified using the universal fish primers COI-WF1 and COI-WR1 (Ward *et al.,* 2005) and the CR was amplified using the universal CR primers L-PROF (Meyer *et al.,* 1994) and 12SAR-H (Palumbi, 1996). From the sequences generated by the universal CR primers, species-specific primers were generated using PRIMER3 0.4.0 (Rozen and Skaletsky, 1999). This resulted in the following primers for *L. mormyrus*: LMINW-L (5'-GCGTTCTTCGCTTAAACTATCC-3') and LMINW-S (5'-TCTTAGCTTTCGTGGCTTCAG-3') for amplifying 1500bp of the CR incorporating hyper variable regions I and II. Internal primers were designed for sequencing outwards from either side of a central G/C secondary structure separating the hyper variable regions: (LMINT-L: 5'-CCTTAAAGGGTTTGATTATTGAGG-3'; LMINT-S: 5'-TCCAGGGATCATTAACACTTCTG-3'). For *S. salpa* the following primers were designed for amplifying the CR: SINW-L (5'-TCCTCTGAAACACAACCACTC-3') and SINW-S (5'-CTTAGCTTTCGTGGGGTCAG-3') for amplifying 1500bp of the CR incorporating hyper variable regions I and II. Internal primers were designed for sequencing outwards from either side of a central G/C secondary structure separating the hyper variable regions: SINT-L: (5'-GAAAGGGAGGATAAATAAATCTCTAGG-3') and SINT-S: (5'- GCTCCGGAGATCGCTAATAC-3'). Initial screening of the hyper variable region II yielded little polymorphism in both species and was subsequently discarded from further sequencing and analyses. This left a final sequence of 707bp in the CR hyper variable region I for *L. mormyrus* and 651bp for *S. salpa*.

Table 3.1 Sampling strategy for *L. mormyrus*. Values in brackets denote sequences derived from GenBank.

Country	Locality	Sample size	COI	CR	Microsatellites
Angola	Benguela (BEN)	21	14	19	-
	Namibe (NAM)	35	8	25	28
	Flamingo (FLA)	18	6	10	24
	Tombua (TOM)	35	4	27	31
South Africa	False Bay (FAL)	33	15	11	33
	Agulhas Bank (AGU)	130	8	29	37
	Struisbaai (STR)	-	(1)	-	-
	Durban (DUR)	-	(2)	-	-
	KwaZulu-Natal (KZN)	-	(1)	-	-
Portugal	Quarteira (QUA)	21	14	19	-
Turkey	Antalya (ANT)	20	1 (20)	4	-

Table 3.2 Sampling strategy for *S. salpa*. Values in brackets denote sequences derived from GenBank.

Country	Locality	Sample size	COI	CR	Microsatellites
Angola	Lucira (LUC)	11	-	10	-
	Flamingo (FLA)	96	24	30	31
South Africa	Infanta (INF)	30	4	-	25
	Rebelsrus (REB)	47	8	26	31
	Port Elizabeth (PEL)	48	8(1)	11	32
	KwaZulu-Natal (KZN)	-	(4)	-	-
Turkey	Antalya (ANT)	10	7 (21)	14	-

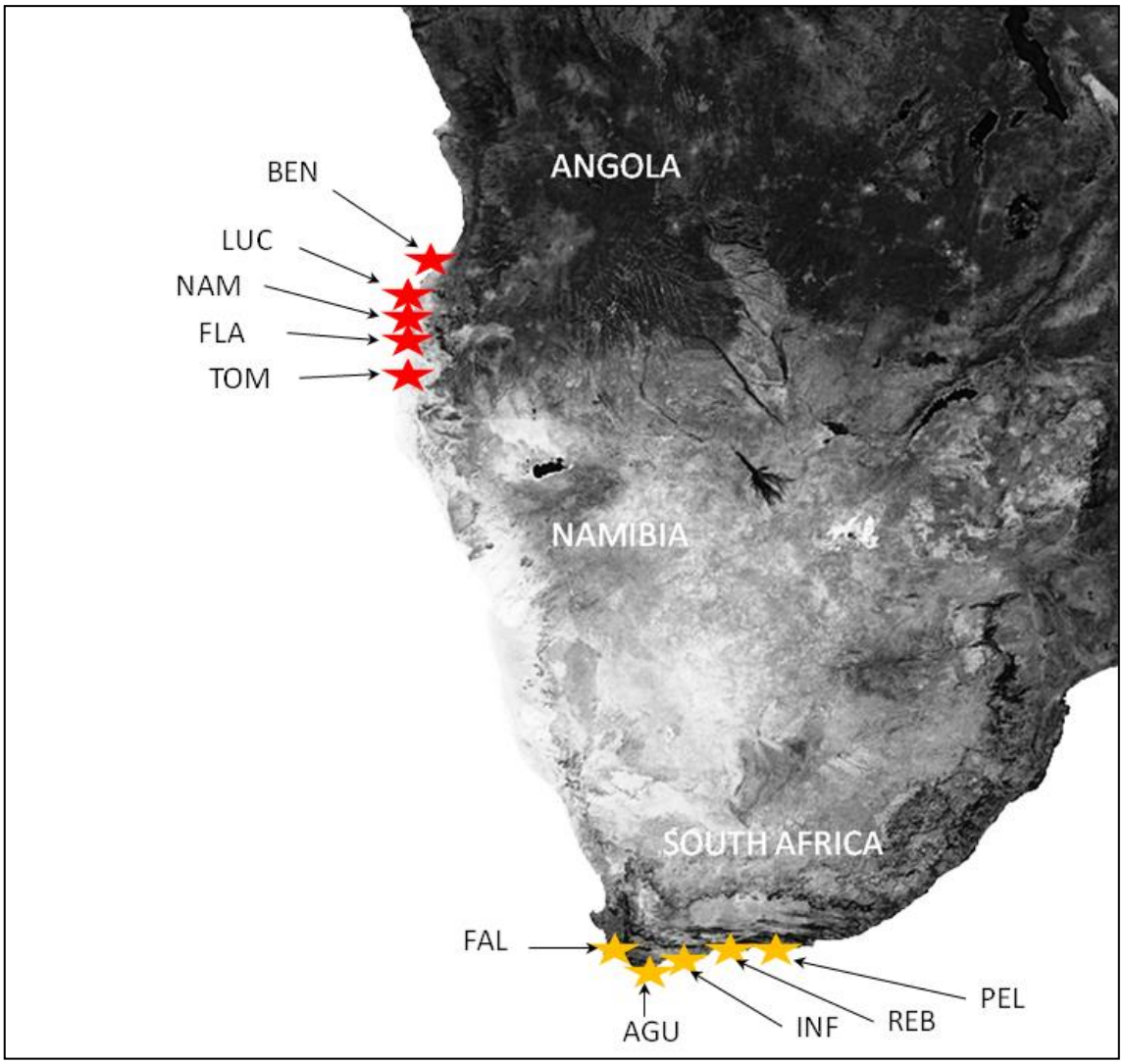

Figure 3.3 Sampling strategy for *L. mormyrus* and *S. salpa.* Sample site ID codes can be found in Tables 3.1 and 3.2.

PCR comprised of the same reaction mix and thermal conditions for both COI and CR markers as previously outlined in the methodology of Chapter 2. The *L. mormyrus* sample from Turkey yielded poorly preserved DNA that in many cases could not be PCR amplified for the targeted mtDNA regions and only 4 CR sequences were obtained.

3.2.3 mtDNA data analysis

CR and COI sequence chromatograms were prepared for analyses as per Chapter 2 methodology. Genetic diversity indices, genetic structure tests (and visualisation) and phylogenetic analyses (network and tree construction) were performed as per Chapter 2. Power Analysis was performed in POWSIM with methods as described in Chapter 2.

In addition to ML phylogenies Bayesian phylogenies were reconstructed using MrBayes 3.2 (Ronquist *et al.*, 2012). Two independent sets of four chains were run for 1 million MCMC generations with parameters recorded every 100th generation. Convergence and stationarity of runs were estimated by examination of likelihood plots and when variance of split frequencies was below 0.01 (Lemey, 2009). If both sets failed to converge after 1 million MCMC generations, the analysis was allowed to continue running until convergence was identified. A consensus tree was obtained after a burn-in period of 2500 generations.

Fu's Fs (Fu, 1997) and Tajima's D (Tajima, 1989) tests were calculated in ARLEQUIN to test for deviations from mutation-drift equilibrium that could be attributed to selection and / or population size changes. To test for signals of past population expansion sequence mismatch distributions and the expansion parameter τ (Rogers and Harpending, 1992), were estimated using ARLEQUIN and DnaSP, with ARLEQUIN providing 95% CI estimates. Additionally, τ was also estimated using SITES (Hey and Wakeley, 1997). SITES uses a Maximum Likelihood method taking into account the distribution of polymorphism in a population and estimates three parameters: θ = ancestral population size, θ = final population size, and τ for a model of sudden population expansion. Given the differing methodologies used to estimate τ some variation is to be expected, however any concordance across these estimates will also help reinforce findings (Kašparová *et al.*, 2015). Estimation of time since population expansion used T = τ/2u (Rogers and Harpending, 1992), where T = time since expansion, τ = expansion coefficient and *u* is the cumulative (across the sequence) probability of substitution. A COI divergence rate of 1.2% per million years was assumed following Bermingham *et al.* (1997), Bowen *et al.* (2001) and Lessios (2008) and 3.6% per million years for CR (Henriques *et al.*, 2014).

NOTE: GenBank sourced COI sequences were included for all regionally 'pooled' analyses for genetic diversity estimates and Φ_{ST} comparisons, with such analyses marked in the relevant results tables. As mentioned above due to the difficulties in amplifying COI from the Turkish *L. mormyrus* individuals in the present study all analyses used Turkish *L. mormyru*s sequences derived from GenBank.

3.2.4 Comparison to other genetic studies on *L. mormyrus* and *S. salpa.*

The Sala- Bozano *et al*. (2009) study utilised CR for *L. mormyrus*, and whilst this was a smaller fragment of CR than utilised in the present study it allows for the direct comparison of data from the present study with that from Sala-Bozano *et al*. (2009). CR sequences from the Sala-Bozano *et al*. (2009) study were obtained from GenBank. Entries contained each single haplotype identified by Sala- Bozano *et al*. (2009), thus lacking the number of individuals with each haplotype and the location of sampling. Regardless of this it is possible to utilise the identified clades in the Sala- Bozano *et al*. (2009) study to infer a coarse level of structuring and comparison to the present study. The CR sequences from the present study and the Sala Bozano *et al*. (2009) study sequences were aligned using BioEdit, identifying a 256 bp region of overlap. Using DnaSP and Network the relationships amongst the resultant haplotypes were reconstructed.

During the writing of the present study a separate study registered on GenBank CR sequence data for *S. salpa*. The study by Paiva *et al*. (unpublished) utilises a 312bp fragment of mtDNA CR and 332 bp of the nuclear S7 ribosomal protein. The Paiva *et al*. (unpublished) study obtained several samples from Portugal, Madeira and within the Mediterranean. These sequences also were in the haplotype format on GenBank, missing the number of individuals with each haplotype and the original sampling locality. For comparison between the present study *S. salpa* CR sequence data and that of the Paiva *et al*. study haplotypes, sequences were aligned in BioEdit using the ClustalW algorithm resulting in 211 bp of overlapping CR sequence for analysis. A median-joining algorithm was used in Network to generate a haplotype network from the resultant alignment.

3.2.5 IMa2 Analysis.

To complement demographic tests IMa2 analysis (Hey and Nielsen, 2007; Hey, 2010), which assigns posterior probability density estimates for population sizes and migration rates from non-recombining sequences using Bayesian Markov Chain Monte Carlo (MCMC) algorithms (Nielsen and Wakely, 2001), was performed. Specifically IMa2 estimates six demographic parameters scaled to the neutral mutation rate inferred for the locus being analysed. All six demographic parameters were estimated for Clade I (shared between Angola and South Africa) for both species. This method was also used to provide an estimate of the time of divergence between Clade I (shared clade) and

Clade II (Angolan endemic) by manually inputting the phylogenetic structure. Other parameters were not inferred in this case as the phylogeography did not conform to a bifurcating model (Sousa *et al*., 2011). Each analysis was run for 1,000,000 burn-in generations and ≥ 5,000,000 sampling generations so that the minimum Effective Sample Size (ESS) across parameters was ≥ 50 (Hey and Nielsen, 2004; 2007). Default and conservative *priors* for splitting time, maximum population size and migration rate, as suggested in the manual were used. We used Metropolis coupling with a geometric heating scheme for one cold chain and 59 heated chains and replicated each run with a different random number seed; the results converged on the same stationary distributions. IMa2 analysis was not performed upon the microsatellite data set for three reasons: firstly the uncertainty surrounding microsatellite mutation rates; secondly the issues of null alleles in the microsatellite data set (see results section) and thirdly due to the rarity of Clade II (only 12) individuals, all of which were not genotyped for the microsatellite component of this study. Initially Clade III (North East Atlantic) was also included in the analyses but no convergence across chains was obtained (ESS values < 5).

3.2.5 Microsatellite DNA markers

Following the testing of 18 published microsatellite loci for consistent PCR amplification and initial screening to assure polymorphism (as outlined in Chapter 2) a subset of five loci for *L. mormyrus* and six loci for *S. salpa* (Table A.1 (Appendix Table 1)) were used for population screening. Samples were amplified using PCR under the following conditions: 300s at 95°C, then 30 cycles of 30s at 92°C, 60s at a primer- and species-specific annealing temperature (Table A.1), 30s at 72°C and a final extension step of 72°C for 120s. All reactions used the following reaction mix: 5 µl of BIOMIX (BioLine), 0.5 pMol of primer (both forward and reverse), 3 µl of template DNA and 1 µl of sterile distilled water giving a total reaction volume of 10µl. Alleles were separated using an AB3730 DNA analyser and allele identity inferred using Peak Scanner 2.

Microsatellite summary and population structure analyses were carried out as per Chapter 2 methodology. Presence / absence of null alleles were also assessed using MICROCHECKER (Van Oosterhout *et al*., 2004) in addition to FreeNA.

3.3 Results

3.3.1 *Lithognathus mormyrus* genetic diversity

The present study identified 175 polymorphic sites across 707bp of mtDNA CR from 144 individuals. Overall genetic diversity in CR for *L. mormyrus* is high: $h = 0.9842$, $\pi = 0.0204$. Regionally (Table 3.3) h is highest in Turkey (1.000 (SD 0.179)), followed by Portugal (0.994 (SD 0.019)), Angola (0.985 (SD 0.006)) and lowest in South Africa (0.958 (SD 0.019)). π is highest in Turkey (0.0320 (SD 0.0215)), followed by Angola (0.0247 (SD 0.0123)), Portugal (0.0135 (SD 0.0072)) and lowest in South Africa (0.0114 (SD 0.0060)). It should be noted that diversity for Portugal and Turkey in particular could be inflated due to the smaller sample size (19 and 4 individuals respectively) compared to Angola and South Africa. Individual sampling site genetic diversity indices are presented in Table A.2.

The present study identified 49 polymorphic sites across 570bp of mtDNA COI from 94 individuals. Overall COI haplotype (h) and nucleotide (π) diversities were 0.674 and 0.0032 respectively. Regionally h (Table 3.4) was highest in Angola (0.723 (SD 0.056)), followed by Portugal (0.725 (SD 0.086)), Turkey (0.638 (SD 0.095)) and lowest in South Africa (0.605 (SD 0.057)). π was regionally (Table 3.4) highest in the Turkey (0.0069 (SD 0.0040)), followed by Portugal (0.0026 (SD 0.0019)), Angola (0.0018 (SD 0.0014)) and lowest in South Africa (0.0014 (SD 0.0012)). Individual sampling site genetic diversities are presented in Table A.3.

Table 3.3 Regional genetic diversity for *L. mormyrus* inferred from mtDNA CR sequences. n: sample size; H: haplotype number; PH: private haplotype number; h: haplotype diversity; π: nucleotide diversity; τ: tau (derived from ARLEQUIN (including 95%CI), DnaSP and SITES); PSSD: the probability that the empirical distribution of mismatches was significantly different from the distribution simulated under a demographic expansion model (P = P-value); Texp: Time since expansion (with 95% CI for ARLEQUIN estimates); D: Tajima's D (P = P-value) and F_s :Fu's FS (P = P-value). Statistically significant values are in **bold.**

	Angola	**South Africa**	**Portugal**	**Turkey**
N	81	40	19	4
H	60	25	18	4
PH	56	21	18	4
***h* (SD)**	0.985 (0.006)	0.958 (0.019)	0.994 (0.019)	1.000 (0.179)
π(SD)	0.0247 (0.0123)	0.0114 (0.0060)	0.0135 (0.0072)	0.0320 (0.0215)
τ (95%CI)	7.748 (2.986; 31.104)	9.074 (4.074;12.980)	2.311 (1.191; 3.621)	-
PSSD (P)	0.0078 (0.389)	**0.0330 (0.03)**	**0.1357 (0.008)**	-
T exp (Arleq)	152 Kya (59 / 610 Kya)	-	-	-
τ (DnaSP)	6.443	2.083	4.155	-
Texp (DnaSP)	127	-	-	-
τ (SITES)	18.6645	3.2554	10.1180	-
Texp (SITES)	367 Kya	-	-	-
***D* (P)**	-0.520 (0.349)	-1.412 (0.055)	-1.003 (0.155)	-
F_S (P)	**-24.074 (0.000)**	**-8.121 (0.008)**	**-8.835 (0.000)**	-

Table 3.4 Regional genetic diversity for *L. mormyrus* mtDNA COI sequences. n: sample size; H: haplotype number; PH private haplotype number; h: haplotype diversity; π: nucleotide diversity. * denotes samples with sequences derived from GenBank.

	Angola	**South Africa**	**Portugal**	**Turkey***
N	32	23	14	21
H	9	4	5	4
PH	6	3	4	4
h (SD)	0.730 (0.056)	0.605 (0.057)	0.725 (0.086)	0.638 (0.095)
π(SD)	0.0018 (0.0014)	0.0014 (0.0012)	0.0026 (0.0019)	0.0069 (0.0040)

3.3.2 *Lithognathus mormyrus* genetic population structure

Φ_{ST} comparisons identified no significant structuring within sampling regions (Angola and South Africa), but generally high and significant Φ_{ST} values for comparisons between sites from Angola and South Africa (Table 3.5), with global Φ_{ST} values of 0.092 (CR) and 0.220 (COI). However whilst this overall pattern is true there are several exceptions when looking at individual sampling site Φ_{ST} comparisons (Table 3.5). The CR Φ_{ST} comparison between Tombua and Flamingo indicated significant structuring, most likely due to the small sample size and the (relatively) small number of private haplotypes and shared haplotypes identified in Flamingo (Table 3.5).

Flamingo also resulted in low and non-significant Φ_{ST} (CR) values with the South African sample sites. False Bay in South Africa also yielded several low and non-significant comparisons (CR) to Angolan sample sites (Benguela, Namibe and Flamingo), again likely due to a relatively small sample size and shared haplotypes between regions. Likewise three COI Φ_{ST} comparisons between South Africa and Angola were high but non-significant (Agulhas Bank / Namibe; Agulhas Bank / Tombua and False Bay / Tombua- Table 3.5): these Φ_{ST} comparisons are likely non-significant due to the lack of private haplotypes in these sites, smaller sample size and the sharing of a common haplotype found in both regions. AMOVA results also indicated significant structuring between Angola and South Africa accounting for 8.70% (CR) and 22.12 % (COI) of the genetic variation in the models (Table 3.6). From a broader regional perspective Φ_{ST} comparison between all regions (including Portugal and Turkey) were high and significant (Tables 3.7).

Table 3.5 Pairwise Φ_{ST} *values* between *L. mormyrus* samples based upon the mtDNA CR (below the diagonal) and COI (above the diagonal) sequence data, between. **Bold** indicates statistically significant Φ_{ST} values. Sample ID codes can be found in Table 3.1.

	BEN	**NAM**	**FLA**	**TOM**	**FAL**	**AGU**	**QUA**	**ANT**
BEN	-	-0.026	0.078	-0.131	**0.24**	**0.198**	**0.574**	**0.912**
NAM	0.006	-	0.13	-0.174	**0.202**	0.163	**0.555**	**0.902**
FLA	0.053	0.017	-	0.076	**0.256**	**0.278**	**0.556**	**0.892**
TOM	0.007	0.007	**0.088**	-	0.156	0.166	**0.536**	**0.892**
FAL	0.025	0.034	0.043	**0.077**	-	-0.074	**0.589**	**0.916**
AGU	**0.149**	**0.118**	0.059	**0.184**	-0.003	-	**0.584**	**0.905**
QUA	**0.385**	**0.412**	**0.525**	**0.411**	**0.485**	**0.565**	-	**0.908**
ANT	**0.799**	**0.845**	**0.884**	**0.814**	**0.863**	**0.917**	**0.889**	-

Table 3.6 Analysis of molecular variance results based on CR and COI sequence data. The AMOVA was structured using Angolan and South African sample regions as groups. **Bold** indicates statistically significant results.

Source of variation	**Percentage of variation CR (P-value)**	**Percentage of variation COI (P-value).**
Among groups	**8.30 (0.000)**	**22.12 (0.001)**
Among populations within groups	2.06 (0.525)	-0.37 (0.525)
Within populations	89.65 (0.065)	78.25 (0.065)

Table 3.7 Pairwise Φ_{ST} values between *L. mormyrus* regional samples based on mtDNA CR (below diagonal) and COI (above diagonal) sequence data. **Bold** indicates statistically significant values.

	South Africa	Angola	Portugal	Turkey
South Africa	-	**0.220**	**0.594**	**0.931**
Angola	**0.092**	-	**0.622**	**0.928**
Portugal	**0.392**	**0.238**	-	**0.908**
Turkey	**0.910**	**0.831**	**0.853**	-

3.3.3 *Lithognathus mormyrus* phylogeography

Reconstruction of the relationships among the identified mtDNA CR haplotypes using a median joining network revealed a clear split between Mediterranean (Turkey) and Atlantic (Portugal, Angola and South Africa) *L. mormyrus* (Figure 3.4). In addition to this major break three other clades are identified utilising the mtDNA CR. Clade I occurs geographically in both Angola and South Africa. Clade II has individuals predominantly occurring in Angola, but also a single individual from False Bay, South Africa. Clade II is rare comparatively to Clade I in Angola (comprising of 13.5% of individuals sampled). Clade III only occurs in the North East Atlantic (i.e. absent from the Mediterranean). Clade I occupies a central position within the network, connected to the other three clades. Following this naming nomenclature, the clade that occurs only in Mediterranean shall be referred to as Clade IV. There are 97 fixed differences between Clades I and IV resulting in a net divergence of 11.67% (SE 1.06%): utilising a divergence rate of 3.6% per Myr yields an estimated divergence time of 3.24 Ma (3.54 - 2.94). Clades I and II are separated by 25 fixed differences with a 2.83% (SE 0.49%) net divergence yielding an estimated divergence time of 786.11 Kya (922.22 - 650.00) Kya. Clades I and III are separated by 13bp, 1.36% (SE 0.35%) net divergence yielding an estimated divergence time of 377.78 Kya (475.00 - 280.56). Full divergence estimates and time since divergence are shown in Table 3.8. In addition Φ_{ST} values were calculated for the three Atlantic clades (Clades I-III): all comparisons were high (Φ_{ST} > 0.48) and significant, with Φ_{ST} values corresponding to the above mentioned levels of divergence (Table A.4).

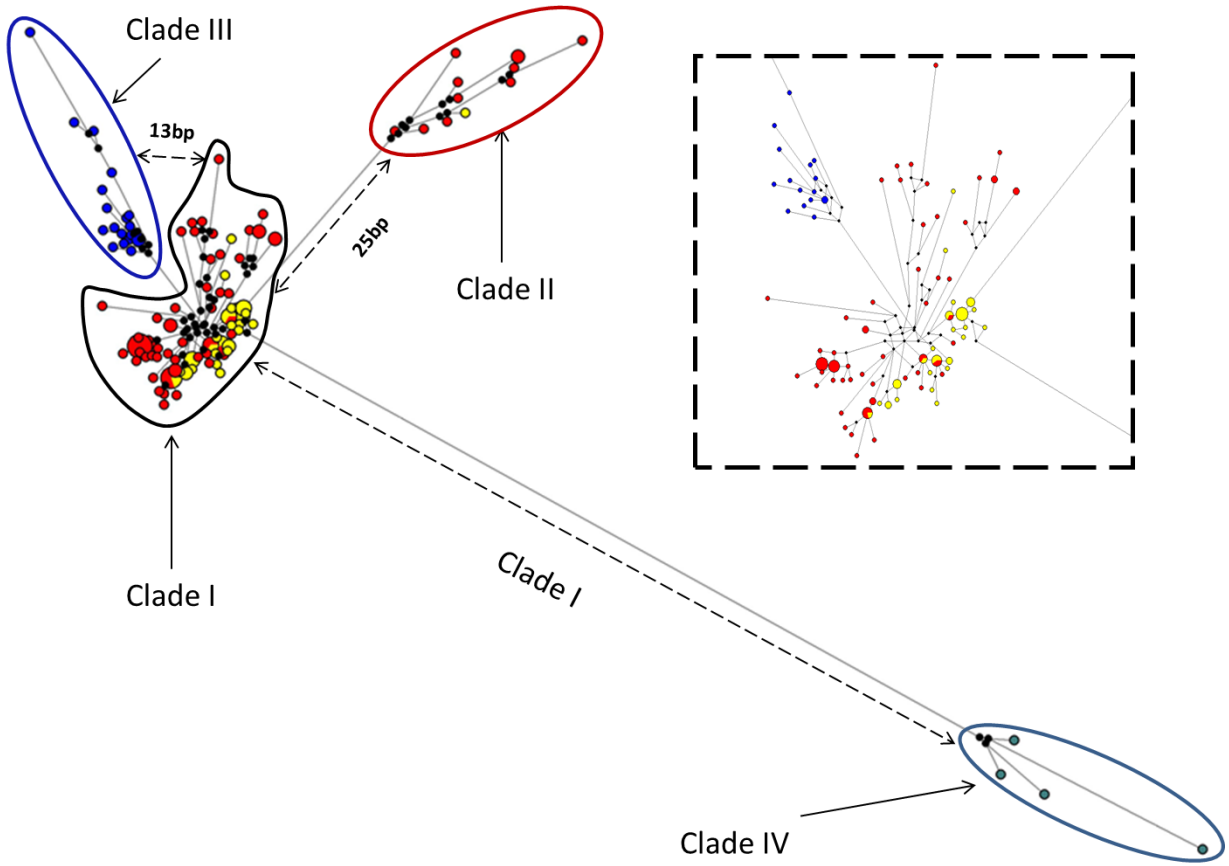

Figure 3.4 Reconstructed median-joining haplotype network for *L. mormyrus*. based on mtDNA CR. Node sizes are proportional to the observed number of individuals bearing that haplotype. Colouring refers to sample regions with Yellow corresponding to South Africa, Red to Angola, Blue to NE Atlantic and Grey to Mediterranean.

Table 3.8 Sequence divergence and times since divergence for *L. mormyrus* CR clades. Below the diagonal estimates of net % sequence divergence (Da) and associated standard errors. Above the diagonal are the estimated times since divergence in millions of years with range in brackets. Values on the diagonal represent mean % intraclade divergence.

	Clade I	**Clade II**	**Clade III**	**Clade IV**
Clade I	1.46 (0.19)	0.79 (0.92/ 0.65)	0.38 (0.48/ 0.28)	3.24 (3.54/ 2.94)
Clade II	2.83 (0.49)	2.16 (0.33)	1.03 (1.20/ 0.88)	3.31 (3.61/ 3.01)
Clade III	1.36 (0.35)	3.74 (0.57)	1.28 (0.21)	3.28 (3.58/ 2.98)
Clade IV	11.67 (1.06)	11.90 (1.08)	11.81 (1.09)	3.08 (0.44)

The combined Sala- Bozano *et al*. (2009) and present study *L. mormyrus* CR haplotypes produced a consistent alignment, including three shared haplotypes. The haplotype network (Figure 3.5) revealed the above mentioned clades I to IV as well as a second highly divergent clade previously described in the Sala- Bozano *et al*. (2009) study corresponding to the Indian Ocean locality (sampled in Durban, South Africa); in the present study this clade is referred to as 'Clade V'. Of clades I-III surprisingly the Sala-Bozano *et al*. (2009) study contained haplotypes present in each clade. This is perhaps not surprising for clades I and III given their sampling from the NE Atlantic and a few individuals from South Africa, but which were not identified by Sala- Bozano *et al.* (2009) as separate clades, perhaps given the close affinity as revealed in the present study. Of most interest is the solitary individual haplotype from the Sala- Bozano *et al*. study that falls within Clade II, which is almost exclusive to Angola in the present study; this haplotype was found in Durban, South Africa, and corresponds to haplotype '72' in the Sala- Bozano *et al*. study, consisting of two individuals.

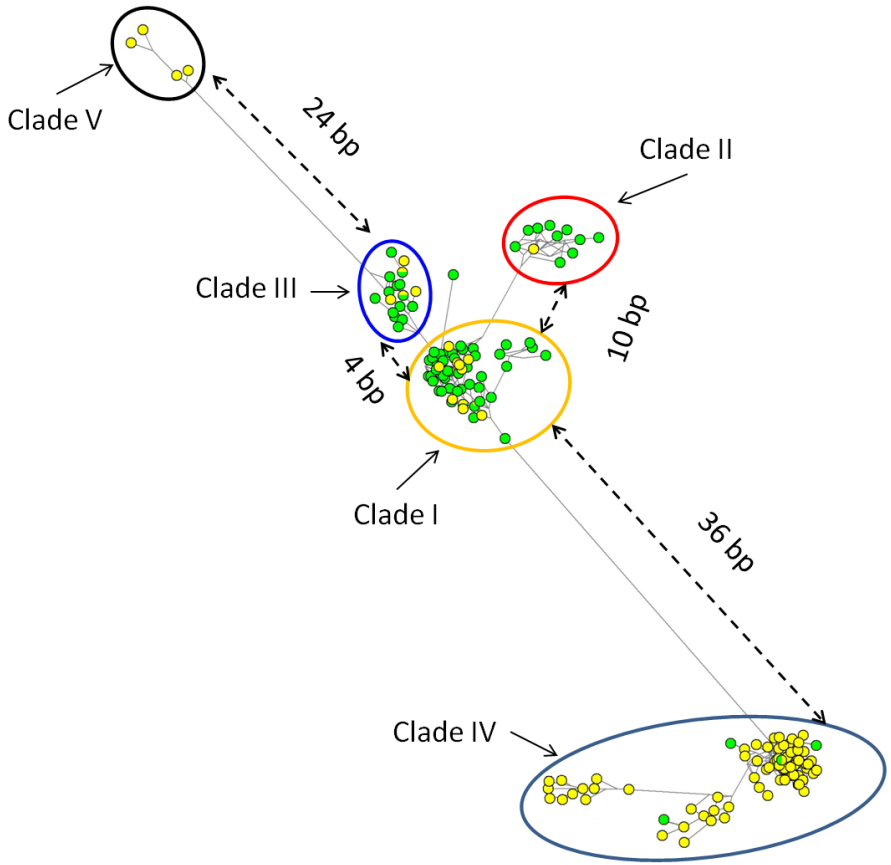

Figure 3.5 Median joining network combining *L. mormyrus* mtDNA CR sequences from the present study and Sala-Bozano *et al.* (2009).The network is based upon 256bp. Branch lengths are equal to mutations steps. Highlighted are the proposed clades presently found in *L. mormyrus*. Note nodes only show the presence of a haplotype and do not reflect the number of individuals carrying that haplotype. Green corresponds to the present study haplotypes and Yellow corresponds to Sala-Bozano *et al*. (2009) haplotypes, with bicoloured haplotypes being identified by both studies. Note median vectors are not displayed to enhance clarity.

Optimal models of nucleotide variation for the *L. mormyrus* CR sequence data set carried out in jModelTest 2.1.4 identified the GTR + G (0.2540) model as best fitting and was used when constructing phylogenies. Reconstruction of the phylogenetic relationships among the identified CR haplotypes revealed a complex pattern. Both the maximum likelihood and Bayesian inferred phylogenies broadly find a similar pattern. Four clades can be resolved from the phylogenies, which correspond to the previously identified clades from the CR network (Figures 3.6 and A.1). Clade I is identified as being a large and diverse basal clade from which the other three clades are derived. Clade II is identified as a monophyletic clade with high support values in both ML (bootstrap branch support value = 895) and Bayesian (posterior probability support value = 0.993) phylogenies. Clade III is also resolved as being monophyletic with moderate support in the ML tree and highly supported in the Bayesian tree (Figures 3.6 and A.1). Clade IV is also identified as being monophyletic and is highly supported by both phylogenies.

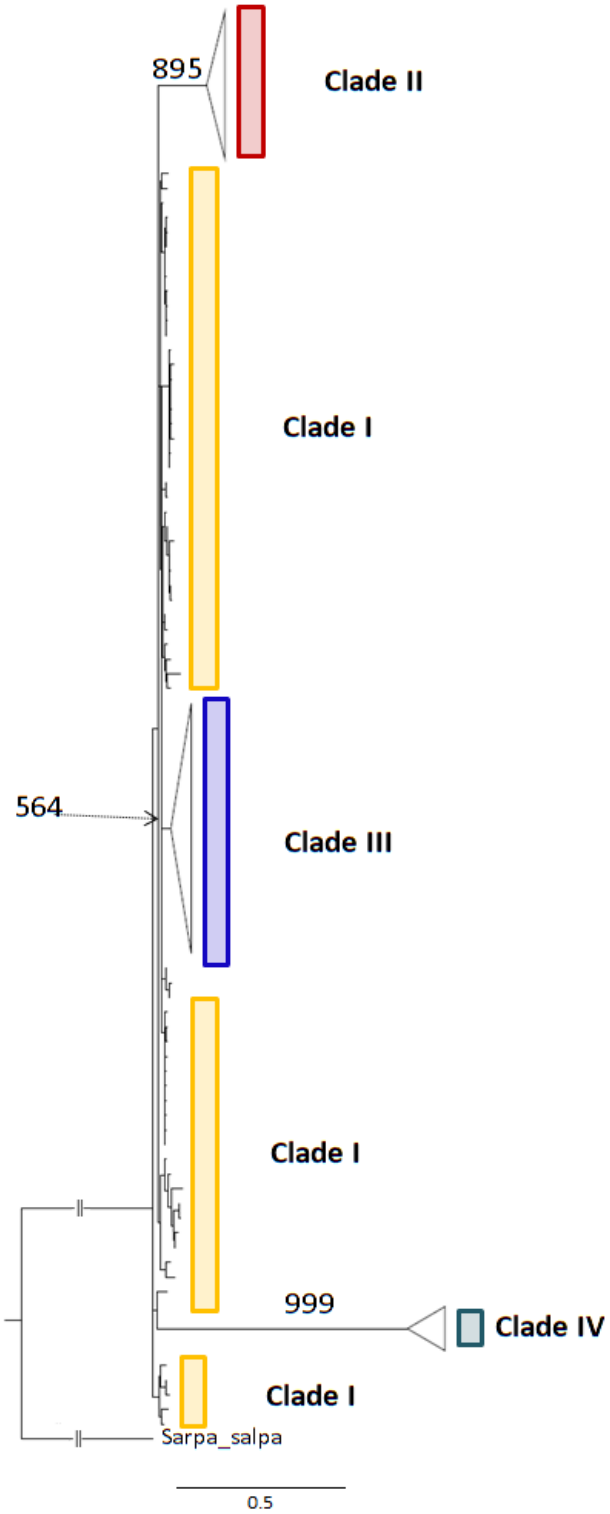

Figure.3.6 Maximum likelihood reconstruction of phylogenetic relationships among CR haplotypes from Angolan, South African, North East Atlantic and Mediterranean populations of *L. mormyrus*. Identified clades (except Clade 1) have been collapsed for clarity and aid interpretation. Bootstrap values are displayed on the branch nodes (1000 bootstraps). Branches to outgroup have been cut (indicated by hashed lines) to aid presentation.

Reconstruction of the relationships amongst the identified COI haplotypes using a median joining network revealed two clear clades which geographically occur in the Mediterranean (Turkey) and Atlantic (Portugal, Angola and South Africa) *L. mormyrus* (Figure 3.7). There are 27 fixed differences between the two clades, resulting in a 4.43% (SE 5.17 - 3.69) net divergence, using a sequence divergence of 1.2% My^{-1} indicates a divergence time of 3.69 Ma (SE 4.31 - 3.08). Within the Atlantic Clade it is possible to identify further structuring (albeit not at the resolution identified by the CR sequence data). The Atlantic clade is dominated by a central common haplotype which occurs in all three regions (South Africa, Angola and Portugal). There is also a putative monophyletic clade composed of all but two of the Portugal COI haplotypes, separated by a single mutation from the common central haplotype; the final Portugal haplotype is a singleton attached to a South African haplotype (Figure 3.7). All other Angolan and South African haplotypes are private to their respective regions and are no more than two mutations from the central common haplotype (Figure 3.7). Clades I-III as identified by the CR sequences are not resolved when using the COI marker. Additionally an outlier singleton haplotype was identified lying between the Atlantic Clade (i.e. clades I-III) and Clade IV; this haplotype corresponds to a GenBank sequence obtained from Durban. It is likely that this haplotype corresponds to the Indian Ocean Clade previously identified in mtDNA CR by Sala-Bozano *et al.* (2009) in *L. mormyrus* samples from the same region.

Optimised models of nucleotide variation for the *L. mormyrus* COI sequence data set carried out in jModelTest 2.1.4 identified the HKY + I (0.7350) model as best fitting and was used in the following phylogenies. Reconstruction of the phylogenetic relationships among the COI haplotypes revealed two divergent clades corresponding to the Atlantic and Mediterranean samples. Both ML and Bayesian phylogenetic trees revealed similar topology with high support values for the identified clades (Figures 3.8 and A.2). Both methodologies identified the Atlantic Clade as being basal and the Mediterranean Clade being derived (Figures 3.8 and A.2). As with the COI network an 'outlier' haplotype ('Clade V') was identified, corresponding to the GenBank sequence from Durban. The placing of the outlier varied between the ML and Bayesian trees. The outlier was found to be the sister taxon to Clade IV (after an initial split between Clades I-III and Clade IV) in the ML phylogeny (Figure 3.8); whilst the Bayesian tree placed it as a sister taxon to both clades I-III and IV (Figure A.2).

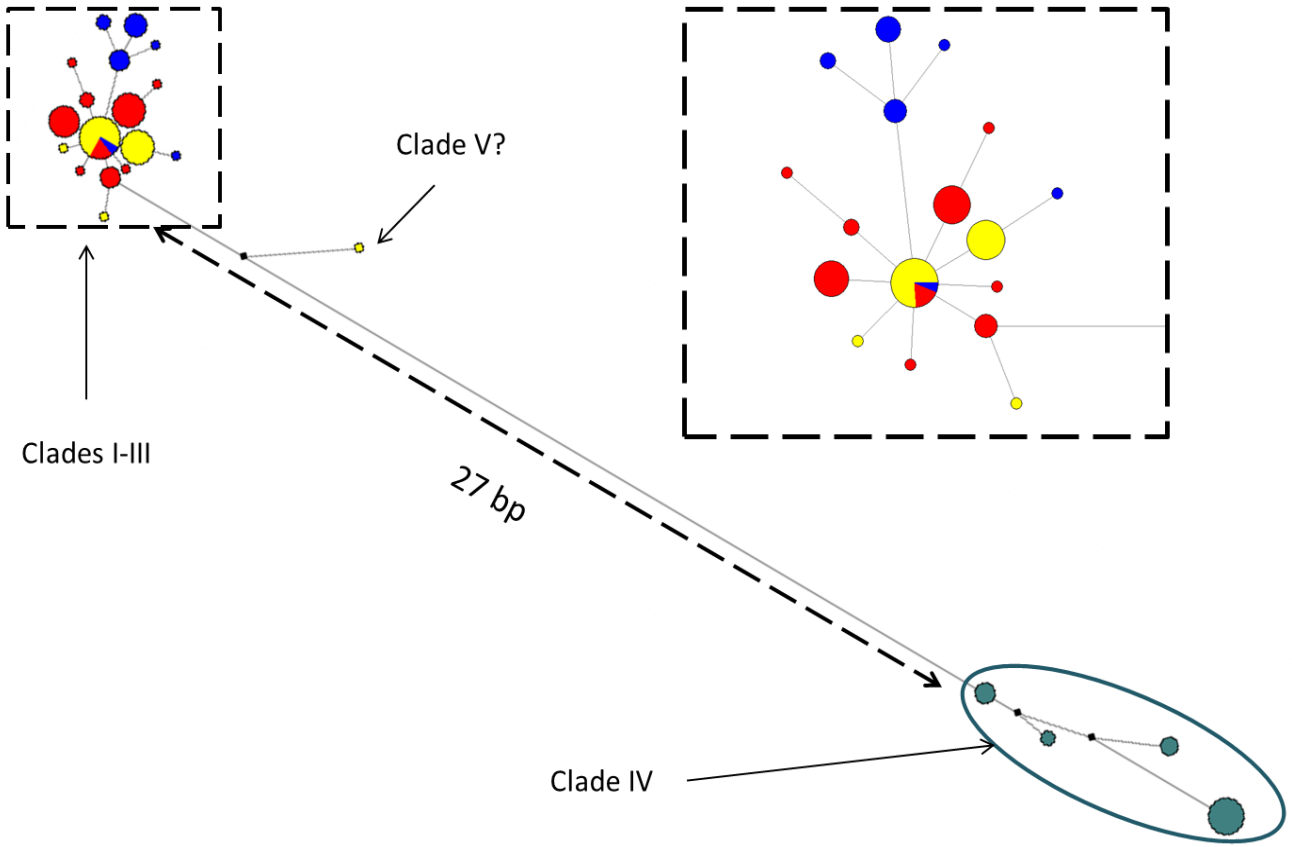

Figure 3.7 Reconstructed median-joining haplotype network for *L. mormyrus* based on mtDNA COI. Node sizes are proportional to the observed number of individuals bearing that haplotype. Colouring refers to sample regions with Yellow corresponding to South Africa, Red to Angola, Blue to NE Atlantic and Grey to Mediterranean.

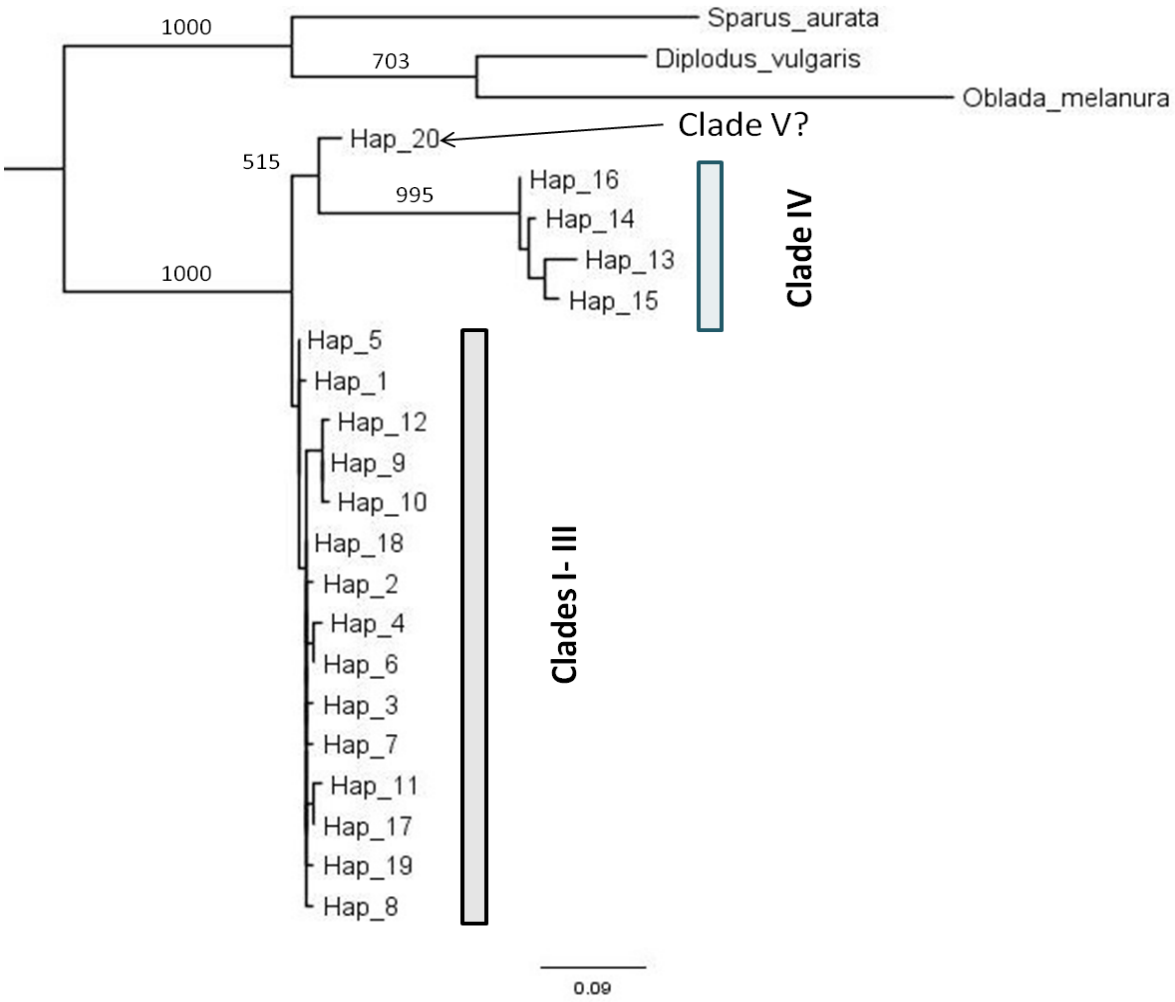

Figure 3.8.Maximum likelihood reconstruction of phylogenetic relationships among COI haplotypes from Angolan, South African, North East Atlantic and Mediterranean populations of L. *mormyrus*. Bootstrap values are displayed on the branch nodes (1000 bootstraps).

3.3.4 Lithognathus mormyrus concatenated COI and CR

Concatenation of the COI and CR sequences was possible for 61 individuals yielding 53 haplotypes and a sequence length of 1322bp. Reconstruction of the resultant haplotype relationships using a median-joining algorithm in NETWORK identified the four previously identified clades (Figure 3.9). This concatenation further allows the identification of the COI haplotype relationships to the CR clades I-III. The common COI haplotype (light blue haplotypes in Figure 3.9) shared between Angola, South Africa and Portugal occupies a central position in Clade I and also occurs in Clade II. In addition Cade II comprises of South African COI (False Bay individual) and an Angolan COI. Clade III is comprised of the putative COI clade identified in Figure 3.7, with the other two Portuguese singleton COI haplotypes occurring as outliers from Clade I (Figure 3.9).

Reconstruction of the phylogenetic relationships of the resultant COI-CR concatenation required the addition of a suitable out group, in this case *S. salpa* COI and CR obtained from the present study. *Sarpa salpa* was chosen due to having comparable sequence length for the CR and being relatively closely related in the Sparid family tree. This addition of an outgroup resulted in no loss of haplotypes and yielded a sequence length of 1318bp. jModelTest 2.1.4 identified the TrN+I (0.3790) + G (0.2550) model as best fitting, however this model is unavailable in PhyML and as such the next best fitting available model identified by jModelTest GTR+I (0.3860) + G (0.2560) was used when constructing the COI-CR phylogenies. Overall the phylogenetic pattern was broadly similar to that identified in the CR phylogenies. Both the Maximum Likelihood (Figure 3.10) and Bayesian (Figure A.3) trees identified five clades with Clade I being identified as the basal taxon from which Clades II and III are resolved as monophyletic clades and Clade IV is again identified as a derived clade.

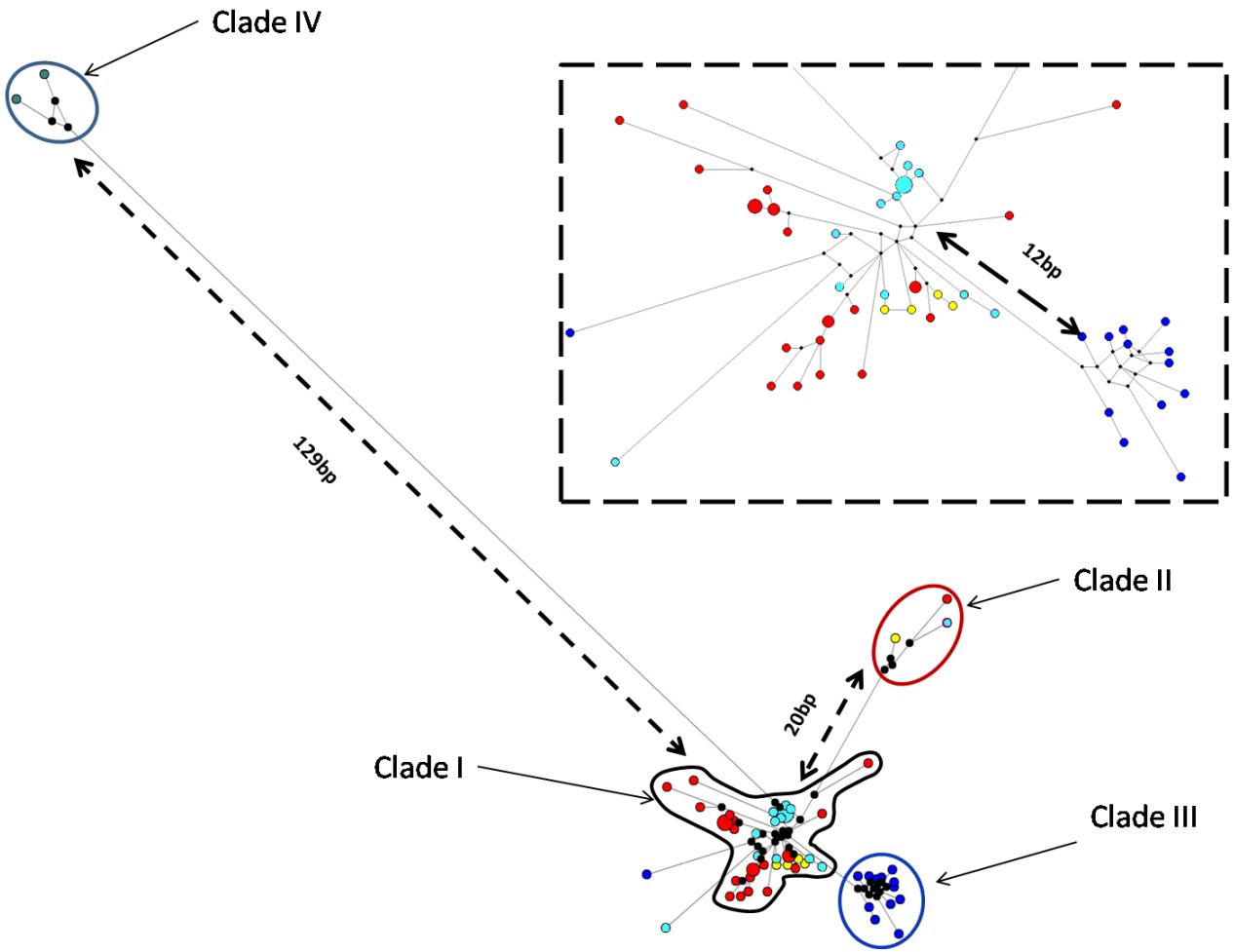

Figure 3.9 Reconstructed median-joining haplotype network for *L. mormyrus* based on concatenated mtDNA COI and CR sequences. Node sizes are proportional to the number of individuals bearing that haplotype with the smallest node corresponding to a single individual. The identified circled clades are the same as those identified using the CR sequences alone and are labelled as such. The haplotype node colouring refers to the COI haplotype geographical origin, as indicated in Figure 3.7: Red =COI haplotypes only occurring in Angola; Yellow =COI haplotypes only occurring in South Africa; Light Blue = the central COI haplotype shared between Angola, South Africa and Portugal; Dark Blue = COI haplotypes that only occurs in NE Atlantic; and Grey =Mediterranean COI haplotypes.

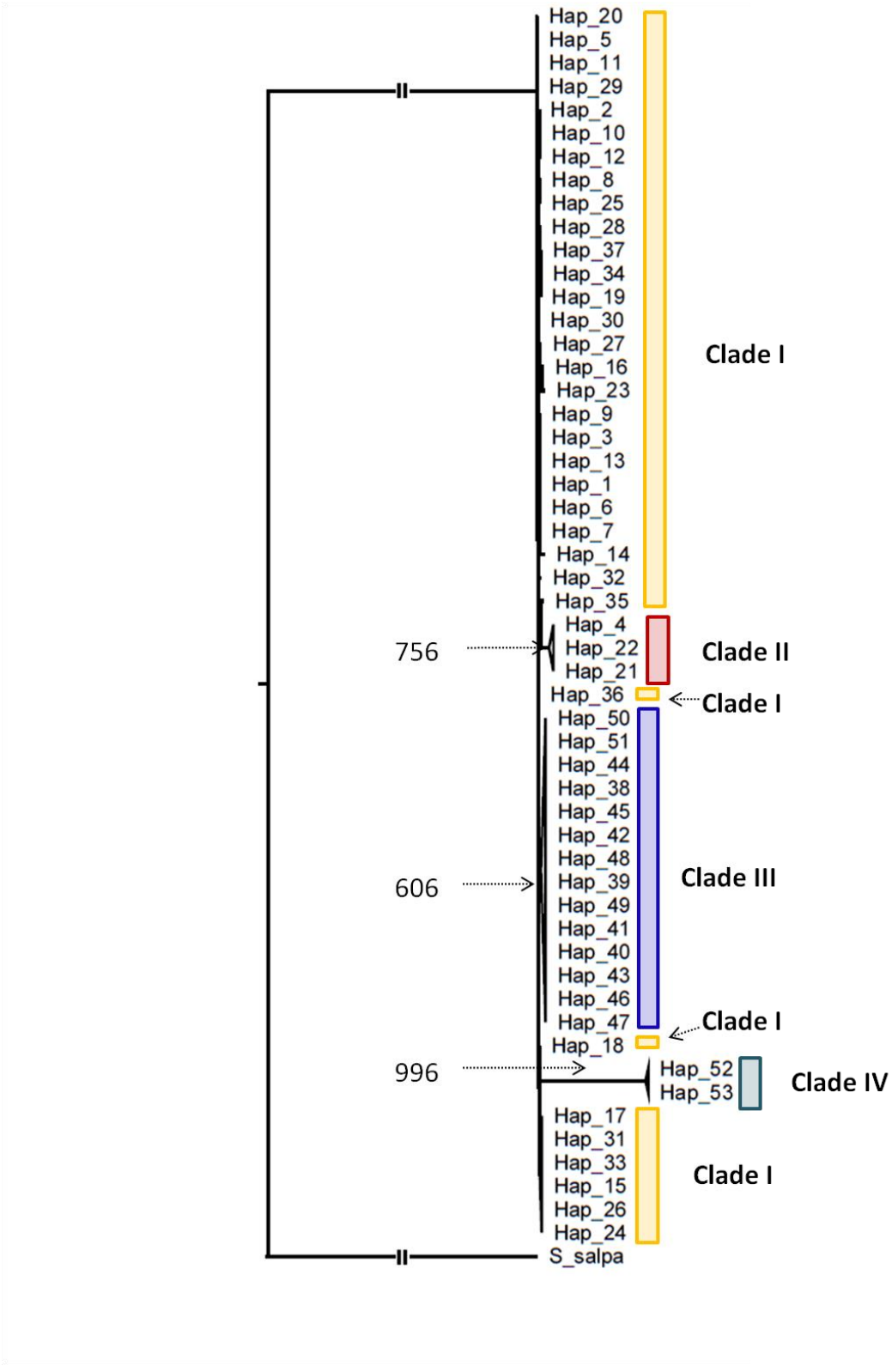

Figure3.10 Maximum likelihood reconstruction of phylogenetic relationships among concatenated COICR haplotypes from Angolan, South African, North East Atlantic and Mediterranean populations of L. *mormyrus*. Identified clades (except Clade 1) have been collapsed for clarity and aid interpretation. Bootstrap values are displayed on the branch nodes (1000 bootstraps). Branches to outgroup have been cut (indicated by hashed lines) to aid presentation.

3.3.5 *Lithognathus mormyrus* demography

Despite high levels of genetic diversity observed in *L. mormyrus* CR, reconstruction of historical demography revealed evidence for past population expansion in many of the studied regions. Demographic analyses were first performed on the geographical regions, but owing to the small number of individuals from Turkey sequenced for CR in the present study it was excluded from the demographic analyses. Tests for neutrality revealed that no region displayed a significant departure from neutrality utilising Tajima's *D*; however all three geographical regions yielded high and significant negative Fu's FS values, indicating a possible population expansion (Table 3.3). However mismatch analyses rejected a scenario of sudden population expansion in North East Atlantic and South Africa but supported a scenario of population expansion in Angola. Three estimates of τ and associated times since expansion were calculated using Arlequin, DnaSP and Sites. Utilising ARLEQUIN τ time since expansion (assuming a divergence of 3.6%/Myr for CR) gave an estimated time since expansion of 152 Kya (95% CI 59 - 610 Kya - Table 3.3). However estimates from the other packages varied, with a more recent expansion time being estimated by DnaSP and a considerably older expansion time estimated by SITES (Table 3.3). Together with the large 95% CI associated with the Arelquin estimate, this casts considerable uncertainty on these results. The most likely candidate for generating such a result is the presence of the two highly diverged clades found in Angola.

Given the observed clade structuring in *L. mormyrus* identified using mtDNA CR, demographic analyses were performed separately on Clades I and II. Analyses were performed on only these clades as Clade III is synonymous with Portugal (and no evidence of a population expansion) and Clade IV synonymous with Turkey which was excluded. Neutrality tests only yielded a negative and significant Fu's FS value for Clade I (Table 3.9). For both clades mismatch analyses could not rule out a hypothesis of past population expansion, with Clade I showing a relatively more recent expansion time of 173 Kya (95% CI 116-305) based on the Arelquin τ= 8.820 (5.920-15.514). Clade II has an older expansion time of 259 Kya (95% CI 180-315) (ARLEQUIN τ = 13.164 (9.166-16.057). Estimates from DnaSP were close to those estimated in Arlequin; whilst SITES consistently identified older dates of expansion (Table 3.9), but the SITES model fitted poorly indicating these results (SITES) should be treated with caution. The discrepancy with SITES estimates is most likely resultant from the

different methodology used to calculate τ using maximum likelihood and incorporating polymorphism distributions within a population (Kašparová *et al.*, 2015). Whilst the DnaSP and ARLEQUIN results are more congruent, DnaSP consistently identifies a more recent expansion time, which could be due to the moment method of Rogers (1995) implemented in DnaSP that is sensitive to deviations from the infinite sites model (Schneider and Excoffier, 1999) making it susceptible to the effects of homoplasy observed in highly polymorphic loci such as CR. Despite these variances between the estimates all three consistently found an older date of expansion in Clade II and a more recent one in Clade I (Table 3.9). Finally Clade I was additionally split geographically to detect any differences in historical demography between Clade I individuals that occur in Angola and South Africa. Tests for neutrality found Tajima's D to not deviate significantly from neutrality in both South Africa and Angola, yet Fu's FS was significant for both regions (Table 3.9). Mismatch distribution analyses did find a difference between the two subsets rejecting a hypothesis of past demographic expansions for individuals of Clade I from South Africa and yet it could not reject a scenario of past demographic expansion for individuals from Clade I in Angola. For comparison dates of expansion were calculated in both subsets yielding near identical dates of expansion: Clade I (South Africa) = 177 Kya (95% CI 96-478) and Clade I (Angola) = 176 Kya (95% CI- 81-254).

Table 3.9 Genetic diversity for clades inferred from *L. mormyrus* mtDNA CR sequences. Clade I is further split geographically. n: sample size; H: haplotype number; PH private haplotype number; h: haplotype diversity; π: nucleotide diversity; τ: tau which is derived from three software packages- ARLEQUIN (including 95%CI), DnaSP and SITES; PSSD: the probability that the empirical distribution of mismatches was significantly different than the distribution simulated under a demographic expansion model (P= P- value); Texp: Time since expansion (with 95% CI for ARLEQUIN estimates; D: Tajima's D (P= P- value) and F_s :Fu's FS (P= P- value). Statistically significant values are in **bold**

	Clade I	**Clade II**	**Clade III**	**Clade I: South Africa**	**Clade I: Angola**
N	109	12	19	39	70
H	70	11	18	24	50
PH	70	11	18	20	46
h (SD)	0.984 (0.004)	0.985 (0.040)	0.994 (0.019)	0.956 (0.020)	0.980 (0.008)
π (SD)	0.0149 (0.0076)	0.0173 (0.0095)	0.0135 (0.0072)	0.0095 (0.0051)	0.0164 (0.0083)
τ (95%CI)	8.820 (5.920; 15.514)	13.164 (9.166; 16.057)	2.311 (1.191; 3.621)	8.977 (4.129; 12.941)	8.988 (4.867; 24.305)
PSSD (P)	0.0059 (0.093)	0.0198 (0.231)	**0.1357 (0.008)**	**0.0456 (0.012)**	0.0061 (0.285)
Texp (Arleq)	173 Kya (116 / 305 Kya);	259 Kya (180 / 315 Kya).	-	-	176 Kya (81 / 254 Kya).
τ (DnaSP)	6.774	11.440	4.155	-	-
Texp (DnaSP)	133 Kya	225 Kya	-	-	-
τ (SITES)	11.154	14.341	10.118	-	-
Texp (SITES)	219 Kya	282 Kya	-		-
***D* (P)**	-1.244 (0.088)	-0.371 (0.385)	-1.003 (0.155)	-0.579 (0.316)	-1.123 (0.109)
***F_S* (P)**	**-24.390 (0.000)**	-2.281 (0.105)	**-8.835 (0.001)**	**-9.019 (0.004)**	**-24.394 (0.000)**

3.3.6 *Lithognathus mormyrus* microsatellites

Individuals from five sampling sites, two from South Africa (False Bay and Agulhas Bank) and three from Angola (Flamingo, Namibe and Tombua), were genotyped for five microsatellite loci. Information on genetic variation for each sample / locus combination is provided in Table 3.10. There were no significant deviations from random associations of genotypes (linkage disequilibrium) detected for any loci pair, either across all samples (data pooled) or in any single sample, indicating that all loci are statistically independent. All loci were variable in each sample with the total number of alleles per locus ranging from 9 (Dvul4) to 39 (DsaMS34) with an average of 21.2. Significant deviations from HWE were found in 22 out of 25 locus/sample comparisons with only Dvul38 for two samples (Flamingo and Namibe) and DsaMS27 in the Tombua sample conforming to HWE (Table 3.10). The presence of null alleles was assessed using MICROCHECKER, which identified the presence of null alleles in all five loci and present in all loci in all samples with the exception of Dvul38 in Namibe and DsaMS27 in Tombua, providing a likely explanation for the deviations from HWE outlined above.

Bayesian clustering in STRUCTURE unanimously supported a model of K=2 according to both log probability (P=1 for K= 2 and zero for other models) and Evanno's delta K interpretation. Clustering of individuals into both clusters followed a clear geographic pattern with one cluster containing all Flamingo, Namibe and Tombua (Angola) individuals while all Agulhas Bank and False Bay (South Africa) individuals assigned to the other cluster using the locprior model (Figure 3.11). However the no locprior model identified some individuals not clustering with their geographical locality of sampling (Figure 3.11). Given the presence of null alleles, F_{ST} values were calculated using FreeNA with F_{ST} comparison values calculated firstly without correcting for null alleles and secondly calculated after correction for presence of null alleles. Significance was assessed after 1000 bootstraps. F_{ST} values were high (>0.02) and significant for all comparisons between Angolan and South African samples but much lower and non-significant for comparisons within regions, both before and after correction for null alleles, albeit with F_{ST} values slightly lower after correction (Table 3.11). Both DAPC runs retained the first 60 principal components accounting for 96.6% of variation. The DAPC run with priors (groups defined as sampling locality) once plotted revealed some separation of the Angolan and South African sampling sites with little overlap

suggesting geographical separation of Angolan and South African *L. mormyrus* (Figure 3.12). However, the *find.clusters* option failed to identify any such clustering pattern (i.e. *find.clusters*= 1).

Table 3.10 Genetic diversity of five microsatellite loci in *L. mormyrus* samples. N- number of individuals genotyped; Na- number of alleles; HE- expected heterozygosity; Ho observed heterozygosity. pHWE- Hardy–Weinberg equilibrium probability Values which as statistically significant are in **bold**. Sample ID codes can be found in Table 3.1.

		NAM	FLA	TOM	FAL	AGU
	N	28	24	29	33	32
	N_a	16	15	14	10	9
Dvul38	**H_E**	0.910	0.925	0.923	0.875	0.859
	H_O	0.821	0.792	0.621	0.303	0.219
	pHWE	0.202	0.204	**0.000**	**0.000**	**0.000**
	N	27	24	30	33	36
	N_a	9	8	9	7	8
DsaMS27	**H_E**	0.867	0.825	0.842	0.764	0.727
	H_O	0.593	0.583	0.767	0.424	0.500
	pHWE	**0.028**	**0.001**	0.544	**0.001**	**0.001**
	N	23	24	23	25	13
	Na	16	21	17	19	5
DsaMS34	**H_E**	0.893	0.953	0.925	0.950	0.726
	H_O	0.304	0.708	0.522	0.280	0.000
	pHWE	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
	N	18	23	27	30	20
	N_a	8	7	7	6	7
Dvul4	**H_E**	0.844	0.816	0.714	0.722	0.765
	H_O	0.111	0.217	0.259	0.233	0.100
	pHWE	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
	N	23	23	24	31	24
	N_a	15	18	17	16	17
Dvul84	**H_E**	0.924	0.931	0.922	0.918	0.927
	H_O	0.565	0.739	0.708	0.387	0.333
	pHWE	**0.000**	**0.004**	**0.034**	**0.000**	**0.000**

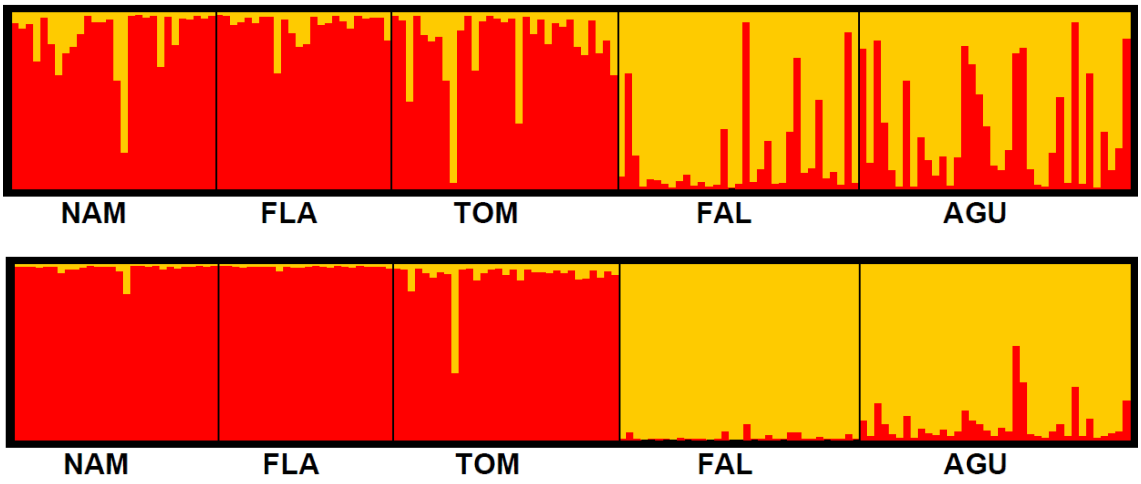

Figure 3.11 Number of genetic clusters observed in *L. mormyrus* populations across the Benguela region. Assignment values for each individual fish obtained from STRUCTURE, based on genotypes from five nuclear microsatellite loci, for K = 2. Cluster 1 (Red) are all Northern (Angolan) population fish; Cluster 2 (Yellow) are all Southern (South African) population fish**.** Top is the plot not assuming priors and bottom is the plot assuming priors. Sample ID codes can be found in Table 3.1

Table 3.11 Pairwise F_{ST} values between *L. mormyrus* samples based on five microsatellite loci calculated in FreeNA, significance assessed after 1000 bootstraps Below the diagonal uncorrected F_{ST}, above the diagonal F_{ST} after correction for null alleles. Significant values are in **bold**. Sample ID codes can be found in Table 3.1.

	NAM	**FLA**	**TOM**	**FAL**	**AGU**
NAM	-	-0.002	0.008	**0.036**	**0.037**
FLA	-0.001	-	0.002	**0.035**	**0.025**
TOM	0.007	0	-	**0.039**	**0.026**
FAL	**0.046**	**0.045**	**0.047**	-	0.013
AGU	**0.049**	**0.039**	**0.035**	0.017	-

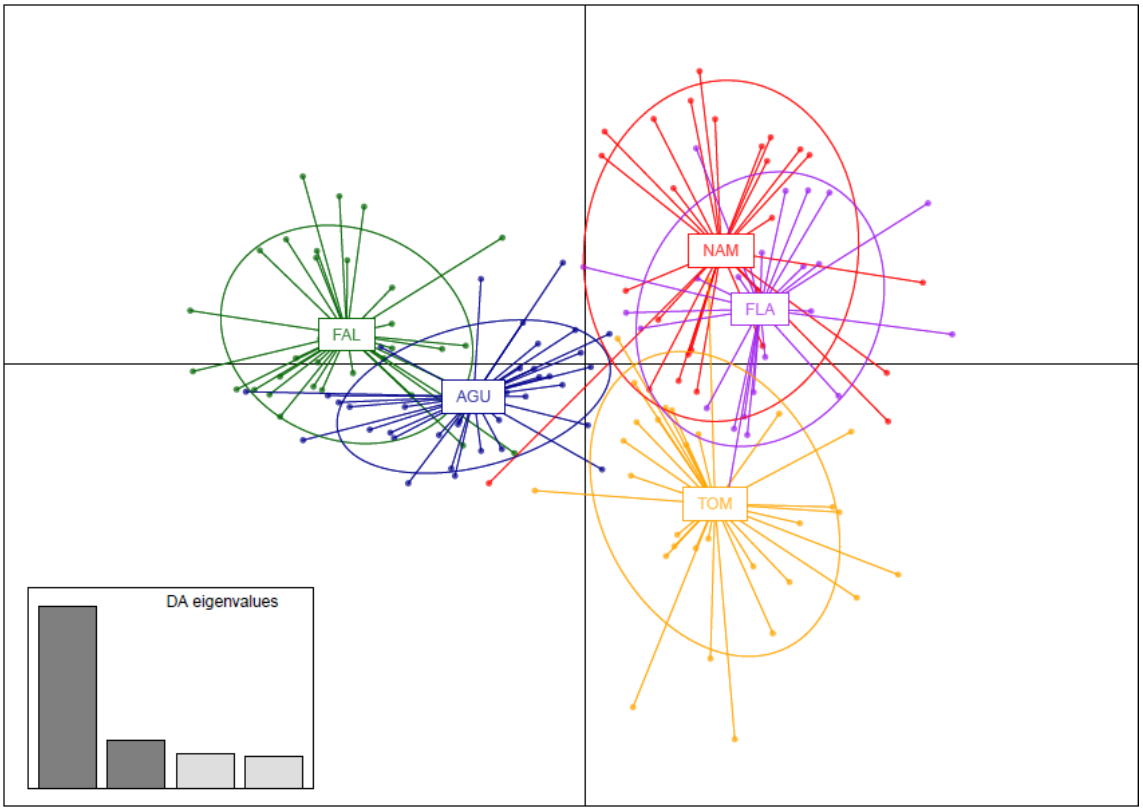

Figure 3.12 Scatter plot of genetic relationships among southern African *L. mormyrus* individuals based on the first two principal components of the DAPC with groups defined *a priori* as per sample site. The graph represents individuals as dots and the groups as inertia ellipses. Eigenvalues of the analysis are displayed in insert. Individual population codes can be identified in Table 3.1.

3.3.7 *Lithognathus mormyrus* power analysis

Power analysis using POWSIM identified that the mtDNA markers presented a low Type I error probability (Fisher P for F_{ST} = 0: CR: 0.06; COI: 0.045), but that the CR had a higher power to detect low levels of differentiation F_{ST} = 0.010 (Fisher P: CR: 0.686; COI: 0.150) and power reached the 95% threshold when F_{ST} = 0.02 (Fisher P: 0.984) whilst for COI the threshold is only met when F_{ST}= 0.08 (Fisher P: 0.970).The *L. mormyrus* microsatellite data had a low Type I error (for F_{ST} = 0; Fisher P = 0.046) and a considerably higher power for detecting between-sample differentiation than the mtDNA markers, with a Fisher P = 0.998 for detecting F_{ST} = 0.0075.

3.3.8 *Sarpa salpa* genetic diversity

Overall genetic diversity in mtDNA for *S. salpa* is high: CR *h*= 0.981 and π = 0.0187; COI *h*= 0.484 and π = 0.0032. Regionally (Tables 3.12 and 3.13) *h* is highest in Turkey (CR = 1.000 (SD 0.177); COI = 0.798 (SD 0.039)), followed by Angola (CR = 0.985 (SD 0.011); COI = 0.554 (SD 0.053)) and lowest in South Africa (CR = 0.958 (SD 0.020); CO = 0.100 (SD 0.088)). π was highest in Turkey (CR = 0.0296 (SD 0.0157); COI = 0.0053 (SD 0.0032)) followed by Angola (CR = 00.0209 (SD 0.0107); COI = 0.0010 (SD 0.0009)) and lowest in South Africa (CR = 0.0055 (SD 0.0032); COI = 0.0002 (SD 0.0003)). It should be noted that the diversities for Turkey could be inflated due to the smaller sample size (14 CR individuals) compared to Angola and South Africa. Individual sample site genetic diversities are reported in Tables A.5 and A.6.

3.3.9 *Sarpa salpa* genetic population structure

Φ_{ST} comparisons using CR and COI sequence data identified a clear pattern of no structuring within sampling regions (Angola and South Africa) and high and significant Φ_{ST} values for comparisons between sites from Angola and South Africa (Table 3.14), with an overall Φ_{ST} value of 0.444 (CR) and 0.304 (COI). However the COI Φ_{ST} comparison between Infanta and Flamingo was lower (Φ_{ST} = 0.160) but not significant. The Infanta and Flamingo Φ_{ST} comparison is likely smaller and non-significant due to Infanta only having a common haplotype which is shared between the two regions: further sequencing of Infanta individuals would most likely identify COI haplotypes private to South Africa as identified in the other two South African sampling localities. Φ_{ST} comparisons between all sampling sites pooled into their regions (including Portugal and Turkey) were high and significant (Table 3.15). AMOVA results also indicated significant structuring between regions accounting for 35.86% (CR) and 35.86% (COI) of the genetic variation in the model (Table 3.16).

Table 3.12 Genetic diversity for *S. salpa* mtDNA CR sequences. n: sample size; H: haplotype number; PH private haplotype number; h: haplotype diversity; π: nucleotide diversity; τ: tau which is derived from three software packages- ARLEQUIN (including 95%CI), DnaSP and SITES; PSSD: the probability that the empirical distribution of mismatches was significantly different than the distribution simulated under a demographic expansion model (P= P- value); Texp: Time since expansion (with 95% CI for ARLEQUIN estimates; D: Tajima's D (P= P- value) and F_s :Fu's FS (P= P- value). Statistically significant values are in **bold.**

	Angola	**South Africa**	**Turkey**
N	40	37	14
H	32	25	14
PH	32	25	14
h (SD)	0.985 (0.011)	0.958 (0.020)	1.000 (0.027)
π(SD)	0.0209 (0.0107)	0.0055 (0.0032)	0.0296 (0.0157)
τ (95%CI)	23.60 (0.55; 38.42)	3.70 (1.90; 4.93	11.85 (7.56; 24.94)
PSSD (P)	0.0266 (0.080)	0.0013 (0.694)	0.0175 (0.247)
Texp (Arleq)	505 Ma (12 / 820 Kya)	79 Kya (41 / 105 Kya)	253 Kya (161 / 530 Kya).
τ (DnaSP)	5.82	3.56	13.04
Texp (DnaSP)	124 Kya	76 Kya	278 Kya
τ (SITES)	0.00	1.97	19.21
Texp (SITES)	Ongoing	42 Kya	410 Kya
D (P)	0.695 (0.816)	**-1.758 (0.018)**	-0.950 (0.168)
F_S (P)	**-11.911 (0.001)**	**-20.263 (0.000)**	**-3.888 (0.031)**

Table 3.13 Genetic diversity for *S. salpa* mtDNA COI sequences. n: sample size; H: haplotype number; PH private haplotype number; h: haplotype diversity; π: nucleotide diversity; τ: tau; PSSD: the probability that the empirical distribution of mismatches was significantly different than the distribution simulated under a demographic expansion model; D: Tajimas D and F_s =Fu's FS. **Bold** indicates significant values.* denotes samples with sequences derived from GenBank.

	Angola	**South Africa**	**Turkey***
N	24	20	28
H	3	2	6
PH	2	9	6
h (SD)	0.554 (0.053)	0.100 (0.088)	0.798 (0.039)
π (SD)	0.0010 (0.0009)	0.0002 (0.0003)	0.0053 (0.0032)

Table 3.14 Pairwise Φ_{ST} values based upon *S. salpa* mtDNA CR sequence data, between *S. salpa* samples. **Bold** indicates statistically significant Φ_{ST} values. Sample ID codes can be found in Table 3.2.

	LUC	**FLA**	**INF**	**REB**	**PEL**	**ANT**
LUC	-	-	-	-	-	-
FLA	0.013	-	0.16	**0.225**	**0.225**	**0.558**
INF	-	-	-	-0.109	-0.109	**0.409**
REB	**0.388**	**0.47**	-	-	-0.143	**0.461**
PEL	**0.346**	**0.441**	-	0.007	-	**0.461**
ANT	**0.494**	**0.538**	-	**0.709**	**0.656**	-

Table 3.15 Pairwise Φ_{ST} values between *S. salpa* regional samples based on mtDNA CR (below diagonal) and COI (above diagonal) sequence data. **Bold** indicates statistically significant values.

	South Africa	Angola	Turkey
South Africa	-	**0.304**	**0.558**
Angola	**0.444**	-	**0.550**
Turkey	**0.747**	**0.536**	-

Table 3.16 Analysis of molecular variance results based on *S. salpa* CR and COI sequence data. The AMOVA was structured using Angolan and South African sample regions as groups. **Bold** indicates statistically significant results.

Source of variation	Percentage of variation CR (P-value)	Percentage of variation COI (P-value).
Among groups	**44.09 (0.000)**	**35.86 (0.012)**
Among populations within groups	0.56 (0.244)	-10.26 (1.000)
Within populations	55.35 (0.335)	74.4 (0.248)

3.3.10 *Sarpa salpa* phylogeography

Reconstruction of the relationships amongst the CR haplotypes using a median joining algorithm in NETWORK revealed high haplotype variation with many unique (singleton) haplotypes but also three clear geographically-associated clades (see Figure 3.13). Clade I contains individuals sampled within both South Africa and Angola, and is the most tightly clustered clade (shortest branch lengths). Clade II contains only individuals sampled in Angola. Clade III contains only individuals from Turkey, and exhibits much longer branch lengths than within the other two clades, with all haplotypes identified being singletons (Figure 3.13). Estimated time since divergence between the three clades yielded the following estimates: Clade I and Clade II diverging ~708 Kya; Clade I and Clade III diverging ~969 Kya; Clade II and Clade III diverging ~852 Kya. Full divergence estimates are outlined in Table 3.17. In addition to the divergence estimates, Φ_{ST} comparisons were performed on the above identified clades with all Φ_{ST} values being large (> 0.64) with higher values being associated with the larger net divergences outlined above (Table A.7).

Table 3.17 Sequence divergence and times since divergence for *S. salpa* CR clades. Below the diagonal estimates of net % sequence divergence (Da) and associated standard errors. Above the diagonal are the estimated times since divergence in millions of years with range in brackets. Values on the diagonal represent mean % intraclade divergence.

	Clade I	Clade II	Clade III
Clade I	0.58 (0.15)	0.71 (0.87/ 0.55)	0.97 (1.14/ 0.80)
Clade II	2.55 (0.58)	1.06 (0.22)	0.85 (1.00 / 0.70)
Clade III	3.49 (0.61)	3.07 (0.54)	2.91 (0.35)

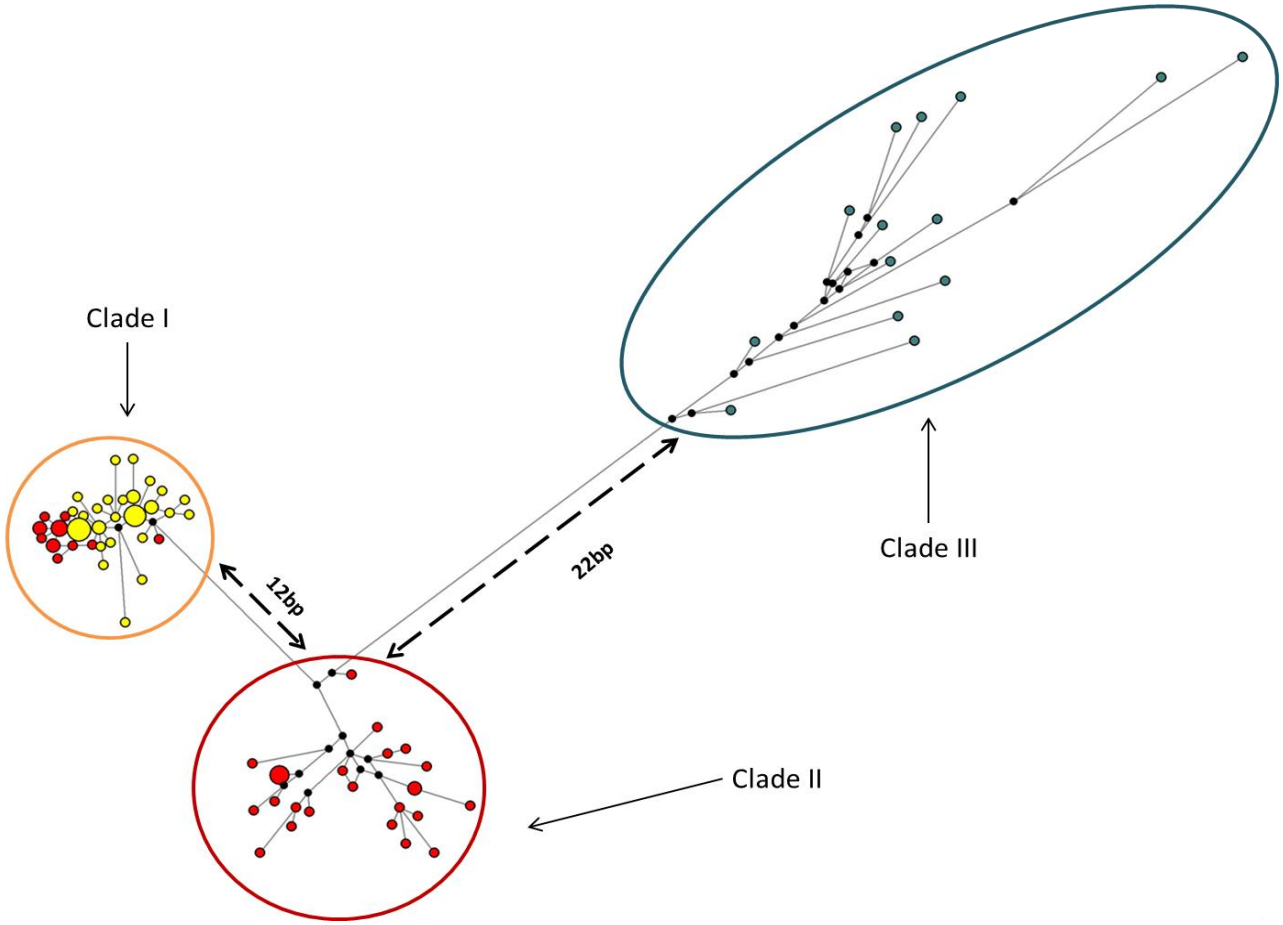

Figure 3.13 Reconstructed median-joining haplotype network for *S. salpa* based on mtDNA CR. Node sizes are proportional to the observed number of individuals bearing that haplotype. Colouring refers to sample regions with yellow corresponding to South Africa, red to Angola and grey to Mediterranean. Observed clades I-III are also encircled.

jModelTest 2.1.4 identified the K80 + I (0.7710) + G (1.7450) model as best fitting and was used when constructing the *S. salpa* CR phylogenies. Reconstruction of phylogenetic relationships amongst the identified CR haplotypes both the maximum likelihood and Bayesian phylogenies identified Clades I- III with moderate to high support (Figures 3.14 and A.4). The maximum likelihood phylogeny identified a weak pattern to the branching order of the three clades, identifying Clade I as ancestral, followed by Clade II and finally Clade III (Figure 3.14). The Bayesian phylogeny however failed to identify any ordering and resolved the relationship amongst the clades as a polytomy (Figure A.4).

Reconstruction of mtDNA COI haplotypes using a median Joining algorithm in NETWORK revealed a single common haplotype present in both Angola and South Africa (Haplotype 2), with two haplotypes private to Angola (Haplotypes 1 and 3) and the singleton haplotype 4 private to South Africa (Figure 3.15). Within Angola Haplotypes 1 and 2 occur at a similar frequency, with Haplotype 1 occurring in 41.67% (10 sampled individuals) and Haplotype 2 two occurring within 54.7 % (13 sampled individuals), with Haplotype 3 being a singleton. The Mediterranean portion of the haplotype network (with individuals only sampled from Turkey) exhibited six haplotypes with four of the haplotypes occurring at similar frequency levels. There is no haplotype sharing between Turkey and southern Africa (i.e. Angola and South Africa; Figure 3.15).

Optimal models of nucleotide variation for the COI sequence data set carried out in jModelTest 2.1.4 identified the HKY + G (0.1840) model as best fitting and was used when constructing phylogenies. Reconstruction of the phylogenetic relationships amongst the identified COI haplotypes revealed little support for any clade structuring within *S. salpa* (Figures 3.16 and A.5). The maximum likelihood tree identified a monophyletic clade with weak support that corresponded to samples from southern Africa (i.e. Angola and South Africa; Figure 3.16), whilst the Bayesian tree failed to identify any such southern African clade (Figure A.5).

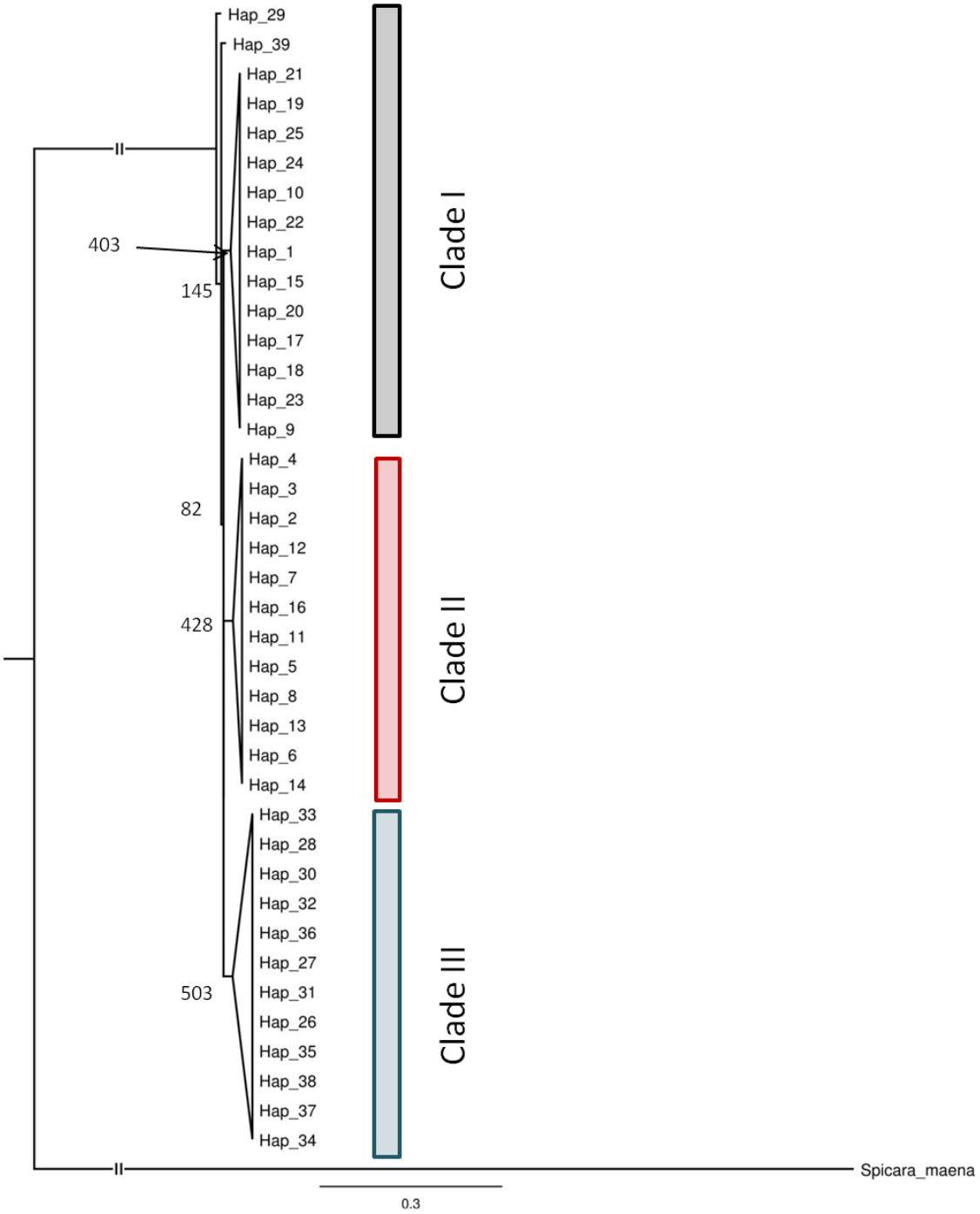

Figure 3.14 Maximum likelihood reconstruction of phylogenetic relationships among CR haplotypes from Angolan, South African and Mediterranean *S. salpa*. Bootstrap values are displayed on the branch (1000 bootstraps). Branches to outgroup have been cut (indicated by hashed lines) to aid display.

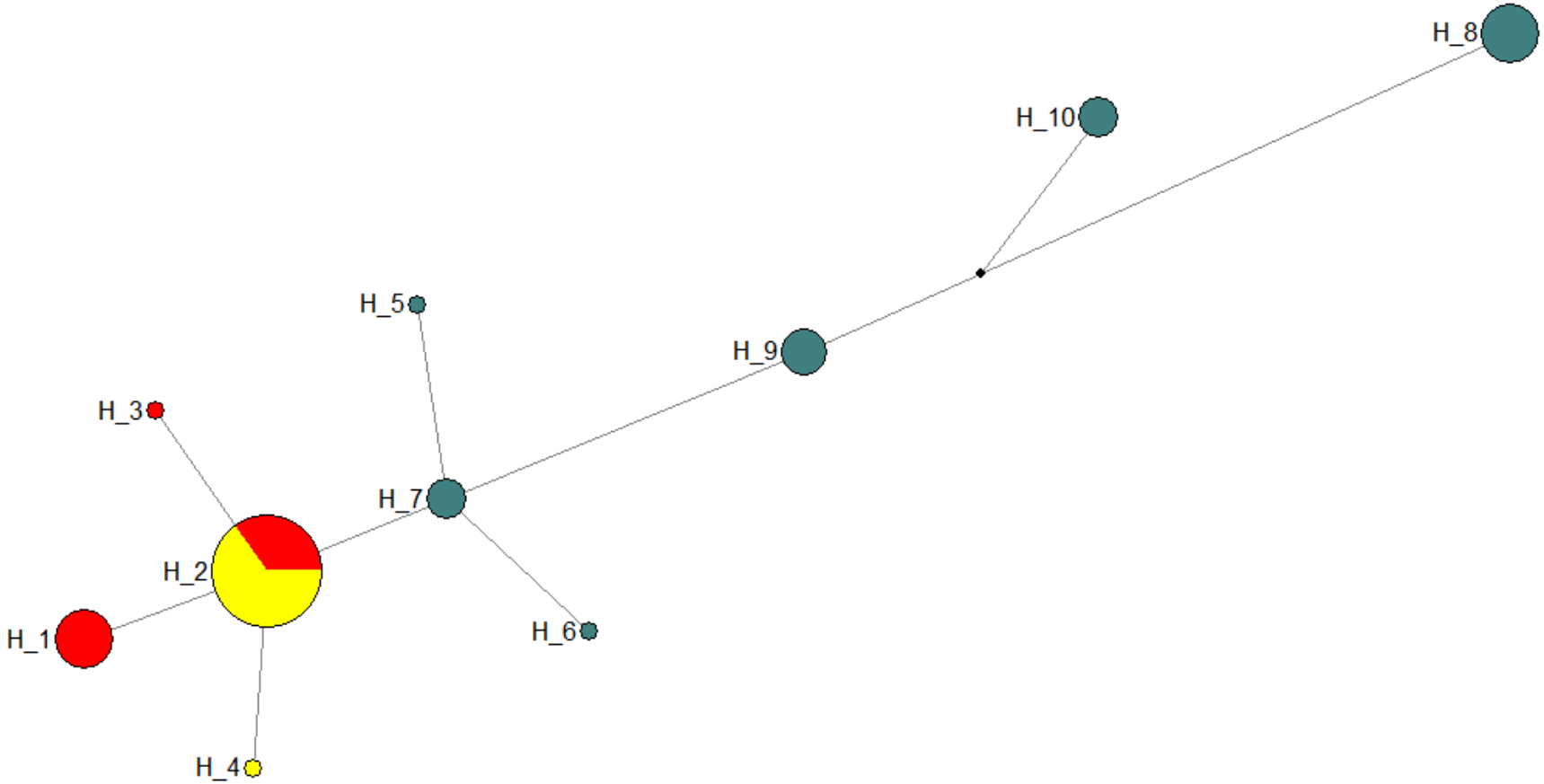

Figure 3.15 Reconstructed median-joining haplotype network for *S. salpa* based on mtDNA COI. Node sizes are proportional to the observed number of individuals bearing that haplotype. Colouring refers to sample regions with Yellow corresponding to South Africa, Red to Angola and Grey to Mediterranean.

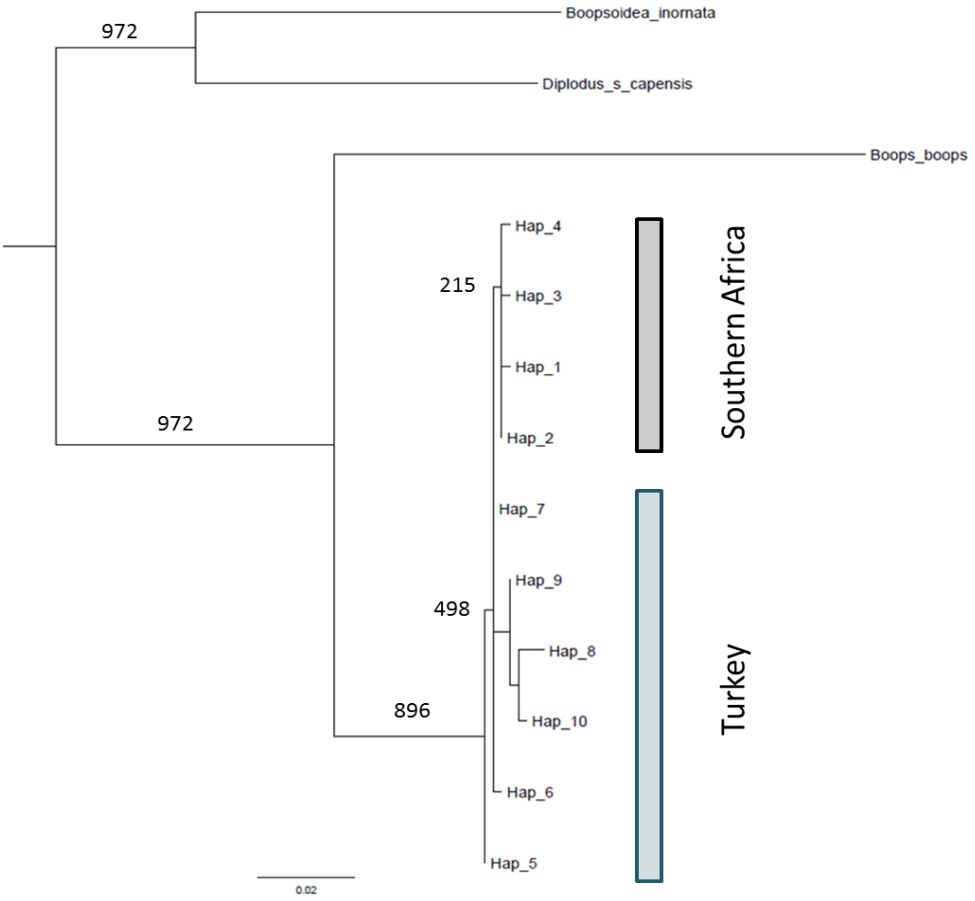

Figure 3.16 Maximum likelihood reconstruction of phylogenetic relationships among COI haplotypes from Southern Africa (Angolan, South African) and Mediterranean (Turkey) S. *salpa*. Bootstrap values are displayed on the branch (1000 bootstraps).

3.3.11 *Sarpa salpa* concatenated COI and CR

The concatenated mtDNA COI and CR network was generated using 32 individuals resulting in 27 haplotypes from a 1252bp long sequence. Reconstruction of the resultant haplotype relationships resolved Clades I to III previously identified utilising the CR (Figure 3.17). The concatenated network also revealed that Clade I (found in both South Africa and Angola) was entirely derived from the common COI haplotype in southern Africa occurring in both Angola and South Africa (Haplotype 2 - Figure 3.15). Clade II (Angolan endemic) was derived from Angolan-only COI haplotypes and the common Angolan / South African COI Haplotype 2 (Figure 3.17). All Mediterranean individuals have COI and CR haplotypes that occur only in the Mediterranean. HKY + I (0.4790) + G (0.0840) was identified by jModelTest model as best fitting and was used when constructing the COICR phylogenies. Phylogenetic trees both ML and Bayesian resolved the three clades both identifying Clade II as being basal and Clades I and III being identified as sister clades (Figures 3.18 and A.6).

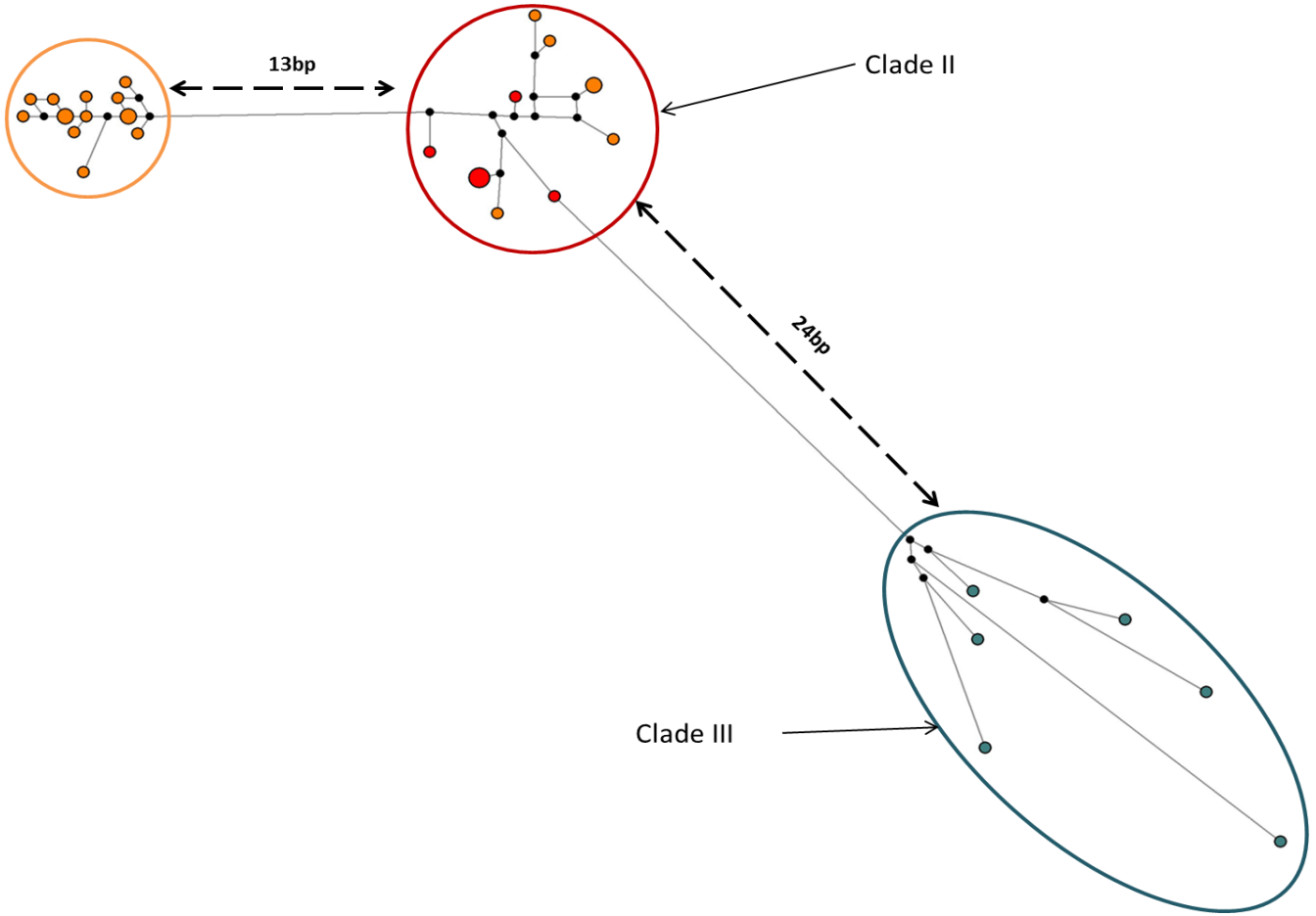

Figure 3.17 Reconstructed median-joining haplotype network for *S. salpa* based on concatenated mtDNA COI and CR sequences. Node sizes are proportional to the observed number of individuals bearing that haplotype, with the smallest node corresponding to a single individual. The identified circled clades are the same as those identified using the CR sequences alone and are labelled as such. The haplotype node colouring refers to the original COI haplotype origin with Red referring to. COI haplotypes only occurring in Angola, Orange haplotypes referring to the common COI haplotype shared between Angola and South Africa and Grey referring to COI haplotypes that occur only in the Mediterranean.

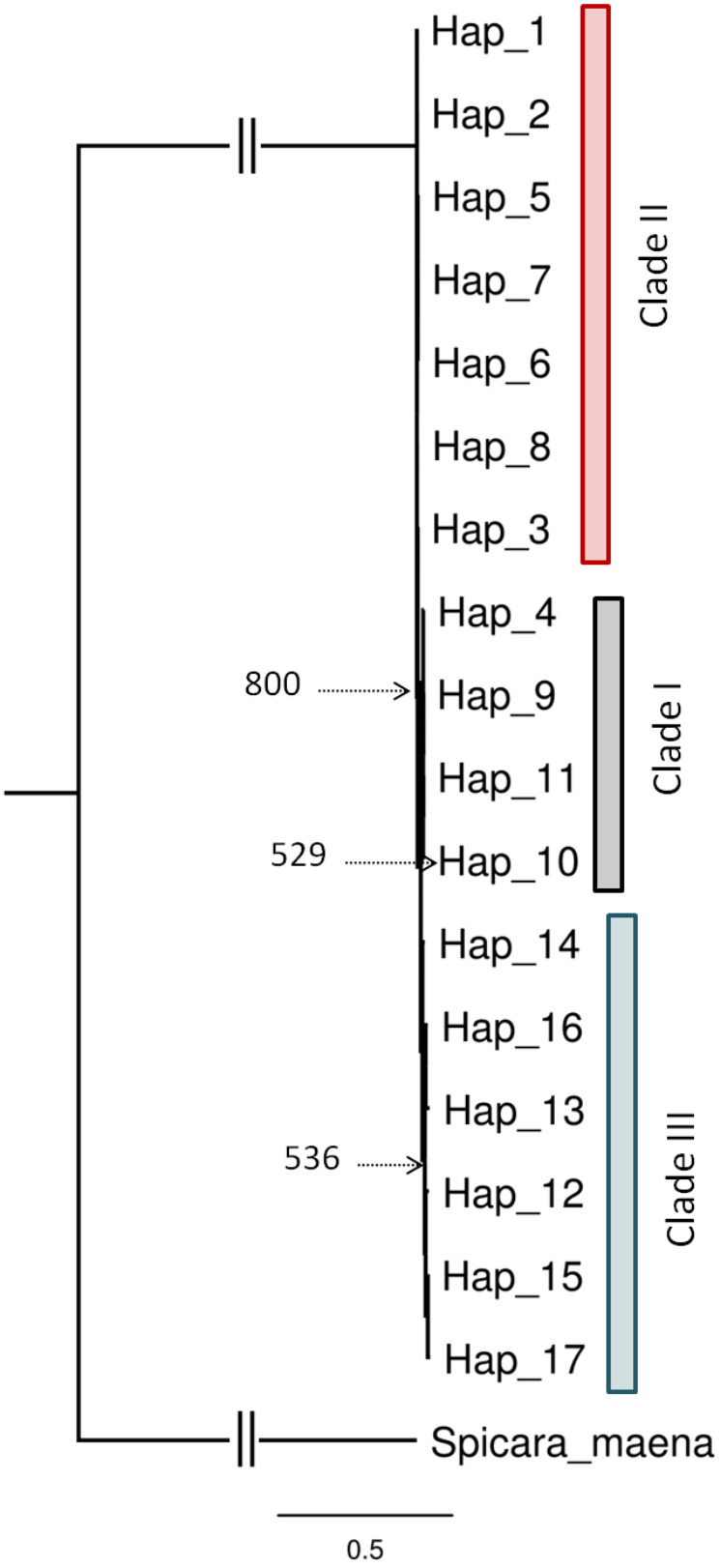

Figure 3.18 Maximum likelihood reconstruction of phylogenetic relationships among the COICR concatenated sequence haplotypes from Angolan, South African and Mediterranean *S. salpa*. Bootstrap values for the clades are highlighted (1000 bootstraps). Branches to outgroup have been cut (indicated by hashed lines) to aid display.

3.3.12 *Sarpa salpa* demography

Despite high levels of genetic diversity observed in *S. salpa,* reconstruction of demographic history using the mtDNA CR revealed evidence for past population expansion. Demographic analyses were first performed on the geographical regions. Tests for neutrality revealed Angolan *S. salpa* to have both significant and negative departures from neutrality in both Tajima's D and Fu's FS, whilst Tajima's D did not significantly depart from neutrality for both South Africa and Turkey, yet Fu's FS displayed negative and significant values in both regions (Table 3.12). Mismatch distribution analyses could not reject a scenario of sudden population expansion in all three regions. As such times since expansion were calculated, with τ being estimated in Arlequin, DnaSP and SITES. Using the τ derived from ARLEQUIN the following time since expansion resulted: South Africa ~79 Kya (95% CI 41-105); Angola ~505 Kya (12-820); and Turkey ~253 Kya (161-530). However estimates from DnaSP and SITES varied considerably (Table 3.12), suggesting (along with the large 95% CI from Arlequin) considerable variation with these estimates. *Sarpa salpa,* similar to *L. mormyrus*, has a relatively complex clade structure which could bias the results, at least for the two southern African regions.

Given the observed CR clade structuring in *S. salpa*, demographic analyses were performed on Clades I and II (with Clade III being synonymous with Turkey and as such results for that clade are presented above). Neutrality tests yielded a negative and significant Tajima's D and Fu's FS value for Clade I, and a negative and significant Fu's FS value for Clade II (Table 3.18). For both clades mismatch analyses could not rule out a hypothesis of past population expansion. Based on Arelquin τ- values Clade I gave a more recent expansion time ~74 Kya (44-115), whilst Clade II gave an older expansion time ~160 Kya (117-194). Estimates of expansion from DnaSP were close to those identified by ARLEQUIN (Table 3.18), whilst SITES estimates of time since expansion identified more recent dates of expansion for both clades. In contrast to *L. mormyrus* the SITES estimates fitted the model well, suggesting the differences observed here in τ estimates are likely due to methodological reasons between SITES and ARLEQUIN / DnaSP outlined above. DnaSP estimates of τ were more closely correlated with the ARLEQUIN estimates for *S. salpa* than they were for *L. mormyrus* (Table 3.18). Again whilst the three estimates varied they were consistent in identifying

a more recent date of expansion in Clade I and an older time since expansion for Clade II (Table 3.18).

Finally Clade I was additionally split geographically in order to detect any differences in historical demography between individuals of that clade which occur in Angola and individuals which occur in South Africa. Tests for neutrality found Tajima's D to deviate significantly from neutrality in Clade I South Africa but not Clade I Angola. Fu's FS was significant for individuals of Clade I from both regions (Table 3.18). Mismatch distribution analyses could not reject a hypothesis of past demographic expansion for individuals of Clade I from both regions. Estimates for timing of population expansion for Clade I (South Africa) were ~80 Kya (38-109) and for Clade I (Angola) ~49 Kya (15-78). This result was found by the other two estimates of time since expansion, albeit with a somewhat more recent date for expansion again being estimated in SITES (Table 3.18).

3.3.13 *Sarpa salpa* - comparison to Paiva *et al.* study

Upon alignment there was 211 bp of overlapping CR sequence for analysis from the present study and that of the Paiva *et al.* CR sequences. A median-joining algorithm was used to generate a haplotype network from the resultant alignment, identifying 151 unique haplotypes, with one haplotype identified in both studies (Figure 3.19). Overall the haplotype network reveals four divergent clades: the three clades previously identified as Clade I (comprising of Angolan and South African individuals), Clade II (Angolan individuals) and Clade III (Mediterranean and North East Atlantic); and an additional clade ('Clade IV'). Both clades I and II are only identified in the present study, whilst clade III is identified by both studies, and Clade IV is exclusive to the Paiva *et al*. study.

Table 3.18 Genetic diversity for clades inferred from *S. salpa* mtDNA CR sequences. Clade I is further split geographically. n: sample size; H: haplotype number; PH private haplotype number; h: haplotype diversity; π: nucleotide diversity; τ: tau which is derived from three software packages- ARLEQUIN (including 95%CI), DnaSP and SITES; PSSD: the probability that the empirical distribution of mismatches was significantly different than the distribution simulated under a demographic expansion model (P= P- value); Texp: Time since expansion (with 95% CI for ARLEQUIN estimates; D: Tajima's D (P= P- value) and F_s :Fu's FS (P= P- value). Statistically significant values are in **bold.**

	Clade I	**Clade II**	**Clade III**	**Clade I - Angola**	**Clade I – South Africa**
N	49	26	14	14	35
H	33	22	14	10	23
PH	33	22	14	10	23
h (SD)	0.972 (0.012)	0.979 (0.021)	1.0000 (0.027)	0.9451 (0.0451)	0.9529 (0.0222)
π(SD)	0.0058 (0.0033)	0.0107 (0.0058)	0.0296 (0.0157)	0.0038 (0.0024)	0.0055 (0.0032)
τ (95%CI)	3.486 (2.049; 5.369)	7.512 (5.492; 9.084)	11.850 (7.555; 24.941)	2.275 (0.721; 3.672)	3.730 (1.770; 5.090)
PSSD (P)	0.0009 (0.714)	0.0139 (0.051)	0.0175 (0.247)	0.0062 (0.622)	0.0015 (0.710)
Texp (Arleq)	74 Kya (44 / 115 Kya)	160 kya (117 / 194 Kya)	253 Kya (161 / 530 Kya)	49 Kya (15 / 78 Kya)	80 Kya (38 / 109 Kya)
τ (DnaSP)	3.756	6.843	13.037	2.207	3.562
Texp (DnaSP)	80 Kya	146 Kya	278 Kya	47 Kya	76 Kya
τ (SITES)	1.262	3.079	19.206	1.641	1.966
Texp (SITES)	27 Kya	67 Kya	410 Kya	35 Kya	42 Kya
***D* (P)**	**-1.665 (0.027)**	-1.032 (0.147)	-0.950 (0.168)	-0.530 (0.331)	**-1.740 (0.023)**
F_S (P)	**-25.995 (0.000)**	**-12.564 (0.000)**	**-3.888 (0.031)**	**-5.587 (0.000)**	**-17.041 (0.000)**

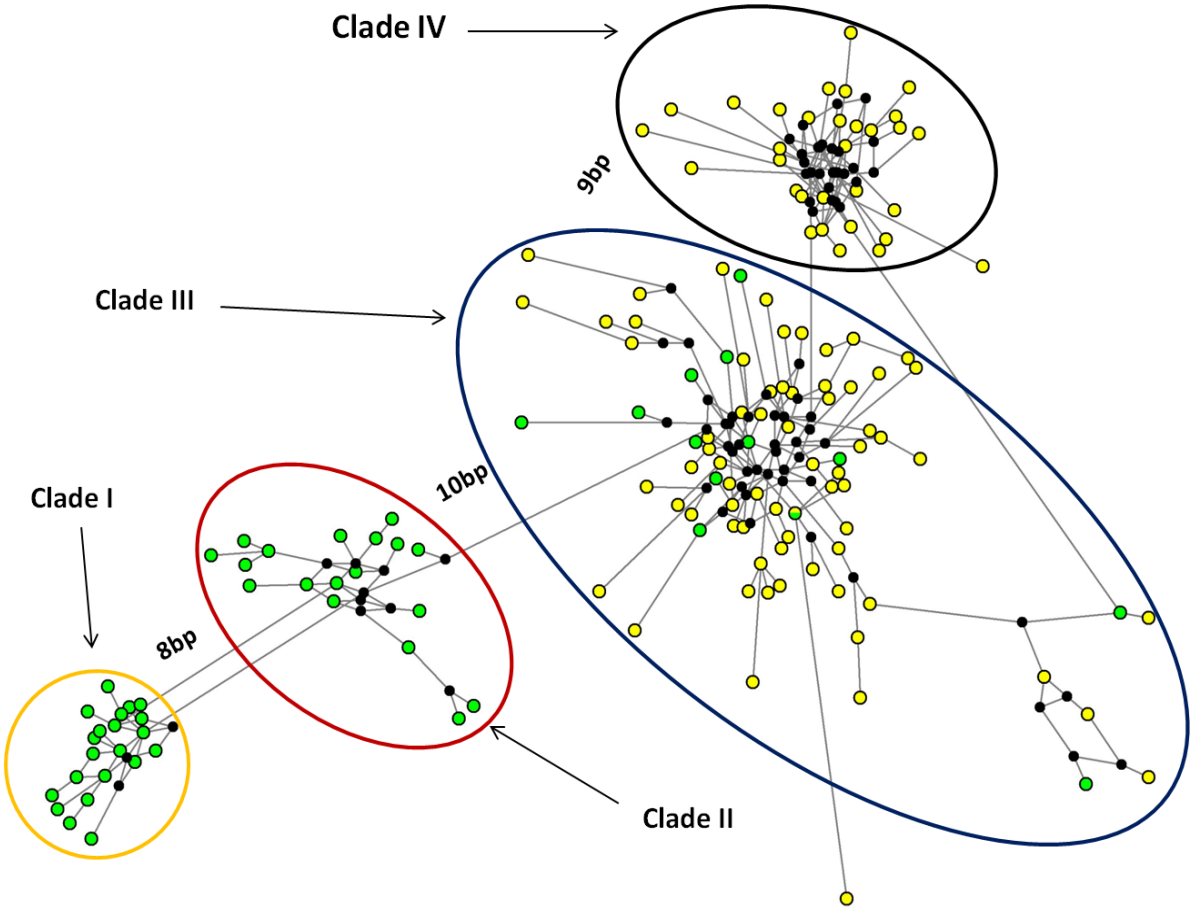

Figure 3.19 Median joining network combining *S. salpa* CR sequences from both the present study and from Paiva *et al*. The network is based upon 211bp. Branch lengths are equal to mutations steps. Highlighted are the proposed clades presently found in *S. salpa*. Geographically clades comprise: Clade I individuals from South Africa and Angola, Clade II individuals from Angola, Clade III comprises from Mediterranean and North East Atlantic and Clade IV of individuals from Madeira and North East Atlantic. Note black nodes only show the presence of a haplotype and do not reflect the number of individuals carrying that haplotype. Green corresponds to the present study haplotypes and Yellow corresponds to Paiva *et al*. haplotypes. Note black nodes represent extinct or unsampled haplotypes.

3.3.14 *Sarpa salpa* microsatellites

Individuals from four sampling sites, three from South Africa (Infanta, Port Elizabeth and Rebelsrus) and one from Angola (Flamingo), were genotyped for six microsatellite loci. Information on genetic variation for each sample/locus combination is provided in Table 3.19. There were no significant deviations from random associations of genotypes (linkage disequilibrium) detected for any loci pair, either across all samples (data pooled) or in any single sample, indicating all loci are unlinked. All loci were variable in each sample with the total number of alleles per locus ranging from three (Dvul33) to 28 (Dvul84) with an average of 19.67 alleles. The presence of null alleles was assessed using MICROCHECKER, which identified the presence of null alleles in four (Dvul38, DsaMS27, Dvul84 and Omel58) of the six loci. Presence of null alleles varied between sampling sites, with Flamingo (Dvul38, DsaMS27 and Dvul84), Infanta (Dvul38 and Omel58), Port Elizabeth (Dvul38, DsaMS27, Dvul84 and Omel58) and Rebelsrus (Dvul38 and Dvul84). Significant deviations from HWE were found in 12 out of 23 locus/sample comparisons (Flamingo 2 of 5 tests; Infanta 3 of 6 tests; Port Elizabeth 4 of 6 tests; Rebelsrus 3 of 6 tests) all due to heterozygote deficits (Table 3.19).

Bayesian clustering supported a model of K=1 with no observable clustering of individuals identified in the STRUCTURE plots (Figure 3.20). Given the presence of null alleles, F_{ST} values were also calculated using FreeNA with values calculated firstly without correcting for null alleles and secondly after correction for presence of null alleles. Significance was assessed after 1000 bootstraps. F_{ST} values were high (>0.01) and significant for the comparisons between Flamingo (Angola) and South African samples Infanta and Rebelsrus both before and after correction for null alleles, albeit with F_{ST} values slightly lower after correction (Table 3.20). However the comparison between Flamingo and Port Elizabeth was non-significant. All F_{ST} values between South African samples were not significantly different to zero. DAPC failed to identify any geographical clustering of individuals with the analysis run with priors (with groups defined as samples) showing overlap of the Angolan Flamingo sample with the three South African samples (Figure 3.21); likewise the *find.clusters* option also did not identify more than one cluster.

Table 3.19 Genetic diversity of six microsatellite loci in *S. salpa* samples. N- number of individuals genotyped; Na- number of alleles; HE- expected heterozygosity; Ho observed heterozygosity. pHWE- Hardy–Weinberg equilibrium probability **Bold** indicates statistically significant values. Sample ID codes can be found in Table 3.2.

		FLA	INF	REB	PEL
Dvul38	**N**	14	20	22	21
	N_a	13	9	12	15
	H_E	0.91	0.841	0.871	0.896
	H_O	0.29	0.6	0.273	0.476
	pHWE	**0**	**0.01**	**0**	**0**
DsaMS27	**N**	23	25	25	32
	N_a	11	14	13	15
	H_E	0.7	0.724	0.726	0.813
	H_O	0.52	0.76	0.72	0.688
	pHWE	0.08	0.855	0.593	**0**
DsaMS34	**N**	29	25	26	32
	N_a	12	14	17	22
	H_E	0.76	0.89	0.878	0.918
	H_O	0.66	0.84	0.769	0.875
	pHWE	0.42	0.109	**0.01**	0.258
Dvul33	**N**	30	24	31	32
	N_a	3	2	2	3
	H_E	0.1	0.042	0.062	0.147
	H_O	0.1	0.041	0.065	0.156
	pHWE	-	1	1	1
Dvul84	**N**	17	14	29	32
	N_a	10	11	18	18
	H_E	0.86	0.796	0.886	0.91
	H_O	0.41	0.643	0.724	0.688
	pHWE	**0**	**0.03**	**0**	**0**
Omel58	**N**	24	12	30	31
	N_a	13	11	15	18
	H_E	0.83	0.813	0.85	0.852
	H_O	0.75	0.583	0.733	0.677
	pHWE	0.1	**0.02**	0.118	**0.01**

Table 3.20 Pairwise F_{ST} values between *S. salpa* samples based on six microsatellite loci calculated in FreeNA, significance assessed after 1000 bootstraps Below the diagonal uncorrected F_{ST} and above the diagonal F_{ST} after correction for Null alleles. Significant values are in **bold.** Sample ID codes can be found in Table 3.2.

	FLA	INF	REB	PEL
FLA	-	**0.028**	**0.031**	0.018
INF	**0.035**	-	-0.023	-0.016
REB	**0.03**	-0.023	-	-0.008
PEL	0.015	-0.016	-0.011	-

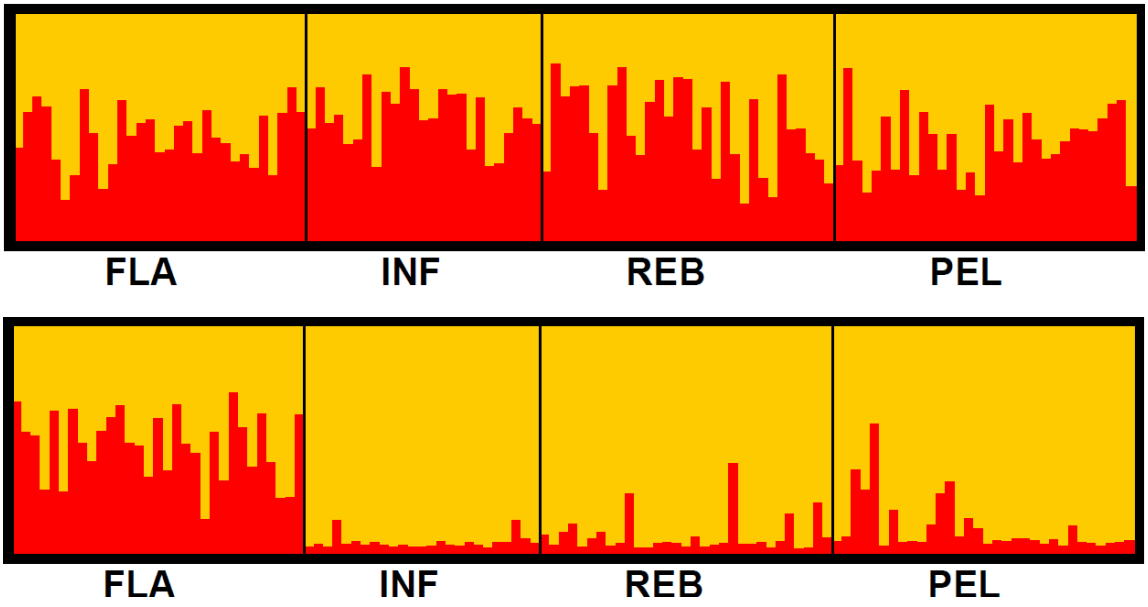

Figure 3.20 Number of genetic clusters observed within *S. salpa* populations across the Benguela region. Assignment values for each individual fish obtained from STRUCTURE, based on genotypes from six nuclear microsatellite loci, for K = 2. Top is the plot not assuming priors and bottom is the plot assuming priors. Population ID codes can be found in Table 3.2.

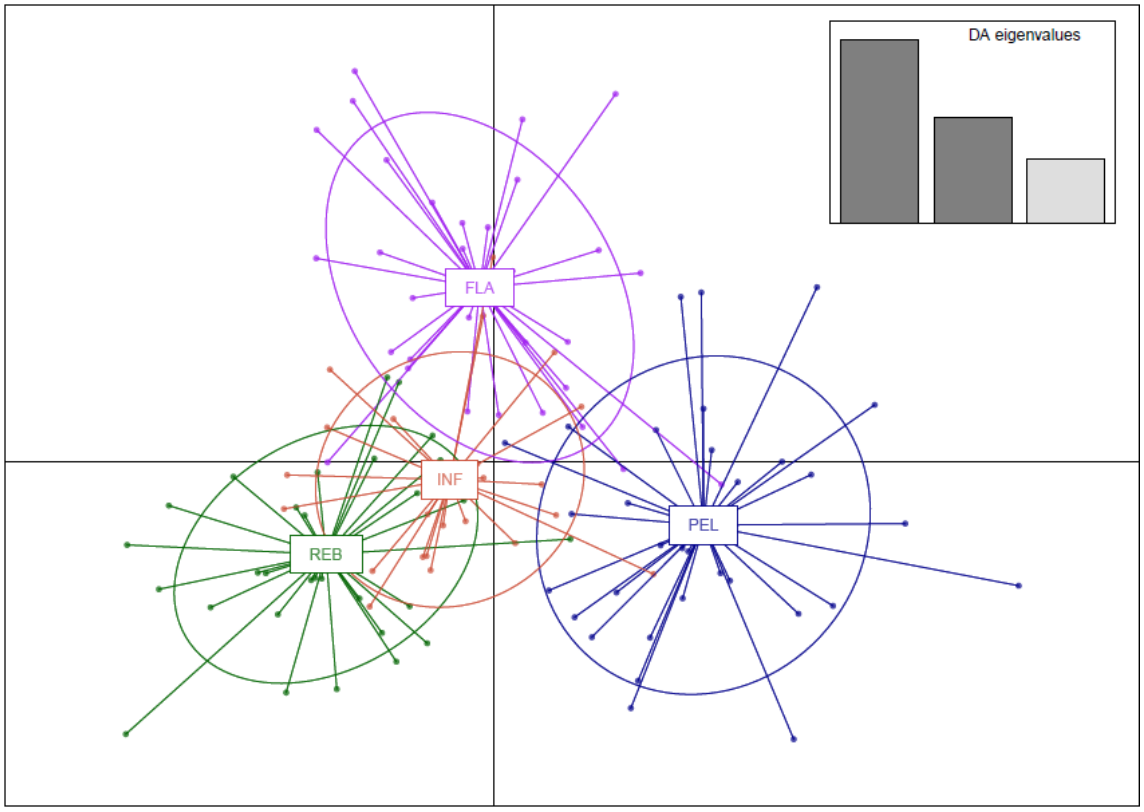

Figure 3.21 Scatter plot of genetic relatedness among southern African *S. salpa* individuals based on the first two principal components of the DAPC with groups defined *a priori* as per sample site. The graph represents individuals as dots and the groups as inertia ellipses. Eigenvalues of the analysis are displayed in insert. Individual population codes can be identified in Table 3.2. .

3.3.15 *Sarpa salpa* power analysis

Both mtDNA markers presented a low Type I error probability(Fisher P for F_{ST}= 0: CR = 0.039, COI = 0.046),but as with *L. mormyrus* the *S. salpa* CR had a higher power to detect differentiation at F_{ST}= 0.010 (Fisher P: CR = 0.470, COI= 0.141) with the 95% threshold met when F_{ST}= 0.025 (Fisher P= 0.951) whilst for COI the threshold is only met when F_{ST}= 0.11 (Fisher P= 0.956).The microsatellite data had a low Type I error(Fisher P for F_{ST}= 0: 0.048) and a considerably higher power compared to mtDNA markers, with Fisher P = 0.984 for detecting F_{ST}= 0.0075.

3.3.16 IMa2 Analysis

In the IMa2 analyses all reported parameters were robustly obtained (ESS > 50). For *L. mormyrus* and *S. salpa* estimates of divergence time between Clade I and Clade II were 588 Kya (95% CI: 390 / 773) and 491 Kya (95% CI: 211 / 838) respectively. As expected times of divergence within Clade I were more recent for both species indicating a second period of vicariance after some intermittent gene flow (Table 3.21 and Figure 3.22). Migration rates were low for both species.

Table 3.21 IMa2 results based upon CR data for both *L. mormyrus* and *S. salpa* showing the present day and ancestral *Ne*, estimated divergence time between Clade I (CI) and Clade II (CII), estimated divergence time for Clade I split geographically (CIA= Clade I Angolan individuals, CIS= Clade I South African individuals) and the directional migration rate south (i.e. from Angola to South Africa) and north (i.e. from South Africa to Angola).

	L. mormyrus	***S. salpa***
N_e Angola	307143	42402
N_e South Africa	150189	133700
Ancestral population size	89583	44950
CI/CII Kya of divergence (95% CI)	588 (390/773)	491 (211/838)
CIA/CIS Kya of divergence (95% CI)	455 (267/745)	262 (59/801)
Migration rate to South Africa	0.0356	0.0553
Migration rate to Angola	0.0152	0.0716

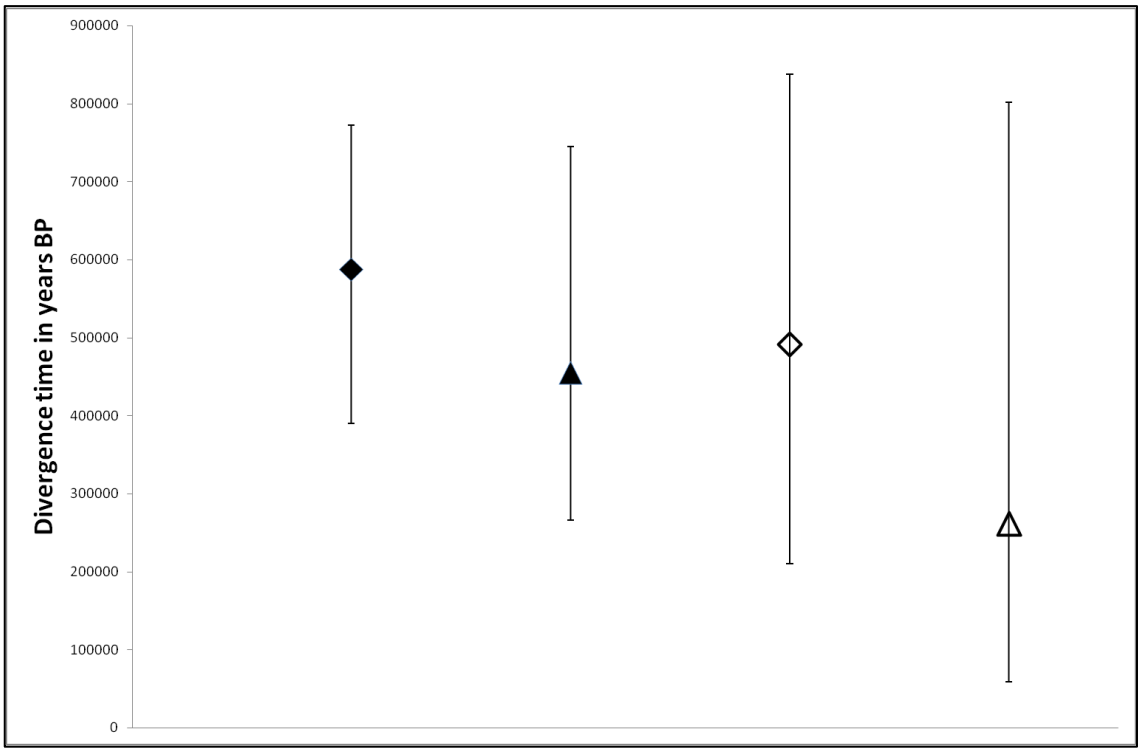

Figure 3.22 Graph showing the different dates of divergence for both *L. mormyrus* (black filled markers) and *S. salpa* (hollow markers). Diamond markers denote the initial divergence between CR Clades I and II. Triangular markers denote divergence between Angolan individuals of Clade I and South African individuals of Clade I. Error bars display the 95% CI associated with the divergence estimates.

3.4 Discussion

While there has been a large number of cross-species comparative phylogeographic studies applied to different marine biogeographic regions (Bowen *et al*., 2016) this is the first study to apply such an approach between closely matched species pairs to the Benguela Current region. Genetic variation across the Benguela system has been studied in a number of marine taxa (Henriques, 2012; Henriques *et al*., 2012; Henriques *et al*., 2014; de Beer, 2014; Henriques *et al*., 2015; Henriques *et al*., 2016; Soekoe, 2016), however in most cases such studies have been largely restricted to the region. By including samples spanning a number of other biogeographical boundaries this study permits a more geographically widespread appraisal of eco-evolutionary divergence. *L. mormyrus* and *S. salpa* revealed congruent phylogeographic patterns consistent with signatures of historical vicariance and genetic divergence in mtDNA associated with biogeographic provinces. Among southern African samples of both species genetic patterns also revealed evidence for asymmetric historical secondary contact across the Benguela Current region, as identified by IMa2 analyses (Table 3.21 and Figure 3.22). However there are also differences in genetic structuring and diversity between the two species, most notably the reduced level of genetic structuring in the *S. salpa* nDNA

microsatellite data across the Benguela Current compared to *L. mormyrus*. IMa2 analyses also identify a more recent date for secondary contact (262 Kya) for *S. salpa* across the Benguela Current than *L. mormyrus* (455 Kya). Such incongruent phylogeographic patterns have often been discarded in comparative phylogeography (Papadopoulou and Knowles, 2016). So whilst the phylogeographic results are broadly congruent there are a number of incongruent patterns between the two species too, which may also further enhance our understanding of phylogeographic processes in the region.

Both species exhibited high levels of mtDNA variability. For *L. mormyrus* CR variation (h= 0.98, Table 3.3) was similar to values reported in previous studies for the species in the Mediterranean Sea (h=0.62–0.90, Bargelloni *et al*., 2003; h=0.92, Sala-Bozano *et al*., 2009). The mtDNA CR exhibited a remarkably high level of singleton haplotypes, with 144 individuals sequenced yielding 97 unique haplotypes, 81 of which were singletons, much higher than in the previous two studies by Bargelloni *et al*. (2003) and Sala Bozano *et al*. (2009). The higher level of singleton haplotypes is likely associated with the longer sequence length used in this study (707bp compared to the 210bp in the Bargelloni *et al*. (2003) study and 263bp in the Sala Bozano *et al*. (2009) study), revealing more polymorphic sites. *Sarpa salpa* CR diversity (h= 0.98) was similar to that in *L. mormyrus* (Tables 3.12 and 3.13) with levels of polymorphism in both species higher than for other coastal fishes studied in the region: *Lichia amia* CR h = 0.867 (Henriques *et al*., 2012); *Atractoscion aequidens* CR h= 0.805 to 0.921 (Henriques *et al*., 2014).

Sarpa salpa and *L. mormyrus* exhibited similar phylogeographic patterns across the sampled range. Within the Atlantic, CR sequences revealed three reciprocally monophyletic clades that were similarly distributed for both species: Clade I occurs in populations immediately to either side of the Benguela Current, but with complete haplotype sorting (i.e. shallow phylogeographic structure with no shared haplotypes between Angola and South Africa) in *S. salpa* whilst *L. mormyrus* revealed incomplete haplotype sorting (i.e. some shared haplotypes); Clade II was restricted to Angolan individuals, except in *L. mormyrus* where a derived haplotype was found in a single South African individual; and Clade III restricted to the North East Atlantic in *L. mormyrus*, but occurs in both the North East Atlantic and Mediterranean in *S. salpa*. However, even with the remarkably similar phylogeographic patterns there are some

differences, notably in the deep Atlantic - Mediterranean divergence in *L. mormyrus* (with a separate monophyletic Clade IV in the Mediterranean), whereas *S. salpa* exhibits no genetic structuring across this area (see below). For *L. mormyrus*, including additional sequence data from Sala-Bozano *et al*. (2009) revealed a fifth highly divergent clade associated with the Indian Ocean and occurring only on the northeast coast of South Africa (Durban). For *S. salpa* including additional sequence data from the unpublished Paiva *et al*. study identified a fourth clade present in the North East Atlantic (Figure 3.19), most likely representing an isolated divergent clade in the Macaronesian island of Madeira as has been identified in other coastal taxa including Sparids (Summerer *et al*., 2001).

3.4.1 Identifying ancestral clades

Reconstruction of the phylogenetic relationships amongst identified mitochondrial clades in *L. mormyrus* identified Clade I as being the basal clade; this was supported in the COI phylogenies (with clades I-III as a single clade), the CR phylogenies and the concatenated COI-CR phylogenies (Figures 3.6, 3.8, 3.10 and A.1- A.3). Such a phylogenetic pattern strongly indicates Clade I as ancestral suggesting *L. mormyrus* originated in the south-eastern Atlantic and then subsequently colonised the North East Atlantic and Mediterranean. An Atlantic origin of *L. mormyrus* is logical since any marine coastal fish that evolved in the Mediterranean would have been extirpated during the Mediterranean salinity crisis 5.33 Ma (Patarnello *et al*., 2007). The estimates of time since divergence also support this view of an Atlantic origin for *L. mormyrus.*

Although the mtDNA phylogenies were shallower in *S. salpa*, the concatenated (COI-CR) tree supports ancestral status for Clade II. The central position of Clade II within the CR network is also consistent with it being ancestral. Therefore, as for *L. mormyrus* the data indicate a south-eastern Atlantic origin for *S. salpa*. The higher levels of genetic diversity observed in *S. salpa* Clade III could be suggestive that it is the ancestral population, but genetic diversity in the southern African clades may have been eroded by more pronounced population crashes during glacial periods (for which there are signals of more recent expansions in these populations), and levels of genetic diversity may be biased by sample size effects (particularly so with the Paiva *et al*. data).

3.4.2 Atlantic / Mediterranean transition

For *L. mormyrus*, the largest phylogenetic divergence was between Clade I (southern African) and Clade IV (Mediterranean), with mean sequence divergence between the clades of 4.43% (COI data). Applying a molecular clock this corresponds to an approximate divergence time of 3.69 Ma (SE 4.31-3.08 Ma). For Control Region sequences the corresponding divergence between clades was 11.67% with an associated time since divergence of 3.24 Ma (SE 3.54-2.94 Ma). Such estimates of divergence are within keeping with the previous divergence estimates from the Sala-Bozano *et al.* (2009) study which estimated the Mediterranean-Atlantic divergence dating to 3.43 Ma (95% CI 1.83–5.03 Ma) and similar to divergence times between the Atlantic and Mediterranean clades identified in other Sparids: *Spondyliosoma* 2.25-2.58 Ma (this thesis and Bargelloni *et al*., 2003); *Dentex dentex* 1.2 Ma (Bargelloni *et al*., 2003). Such dates of divergence are consistent with the known palaeoceanography of the Mediterranean whereby after the Zanclean Flood (5.33 Ma), it was not until the mid-Pliocene that environmental conditions improved enough to allow colonisation of *L. mormyrus* from the Atlantic (Haywood *et al.,* 2000). During subsequent glacial periods (2.58 Ma onwards) many North East Atlantic populations of marine fishes were extirpated or moved south resulting in reduced gene flow between Atlantic and Mediterranean populations (Maggs *et al*., 2008; Miralles *et al*., 2014; Silva *et al.*, 2014.), a scenario which seems likely for *L. mormyrus* also. Upon secondary contact between the two diverged *L. mormyrus* Mediterranean and Atlantic clades in the Alboran Sea, Sala- Bozano *et al*., and (2009) report introgression between the two clades. The present study sampled *L. mormyrus* from Quarteira in southern Portugal (approximately 50 miles west from the Sala- Balzano *et al*. (2009) Atlantic sample but did not identify any Clade IV (Mediterranean) individuals. Likewise the Sala-Bazano *et al.* (2009) study could not identify any Clade III (North East Atlantic) *L. mormyrus* in the main basins of the Mediterranean Sea. Such a result is indicative that introgression between the two clades is largely restricted to within the Alboran Sea.

In contrast to *L. mormyrus, S. salpa* does not exhibit any population structuring across the Atlantic-Mediterranean divide (combined data of this study and Paiva *et al*., unpublished). This finding is also reflected in Figure 3.19 where in the present study *S. salpa* sampled from Antalya (Turkey) clearly cluster within the Atlantic / Mediterranean clade (Clade III in the present study). Bargelloni *et al.* (2003, 2005) suggest that such

contrasting patterns of Atlantic / Mediterranean divergence in Sparid fish may occur across the Atlantic-Mediterranean divide because of differing life history traits and / or recent colonisation events. Whilst both species studied here have similar life history traits *S. salpa* exhibits migratory behaviour in at least parts of its range (Van der Walt and Mann, 1998) whilst *L. mormyrus* are largely sedentary as adults. Such differences in dispersal ability may have affected the chances for the two species to disperse through the Almeria-Oran Frontal system and the Gibraltar Strait with its one-way current flows from the Atlantic to the Mediterranean. However other biogeographic barriers in *S. salpa*'s range (e.g. Benguela Current and Tropical Barrier) do present formidable barriers to *S. salpa* regardless of its greater dispersal potential. Alternatively *S. salpa* may have colonised the Mediterranean relatively recently and as such insufficient time may have passed to accumulate genetic differentiation between the two regions. The relatively old estimates of population expansion (compared to southern African populations) in European *S. salpa* would argue against such a recent invasion, but studies of nuclear genetic variation among North East Atlantic and Mediterranean populations of *S. salpa* will be needed to assess the presence of current gene flow (i.e. the unpublished Paiva *et al*., study).

3.4.3 Divergence of North East Atlantic clades

Both *L. mormyrus* and *S. salpa* exhibit divergent North East Atlantic clades (Clade III). For *L. mormyrus* this clade was less divergent than for *S. salpa*, with corresponding divergence times of 378 Kya and 969 Kya respectively. Furthermore the most closely related clade in *L. mormyrus* to Clade III is not the geographically closest Clade II (as found in *S. salpa*) but the more geographically distant Clade I in South Africa. Collectively this suggests differing dynamics underpinning the divergence of Clade III in each species.

For *S. salpa* the 931 Kya time of divergence is before the first longer and colder ice age after the transition from 41 Kyr to 100 Kyr glacial-interglacial cycles during the Mid-Pleistocene Transition (MPT). Before this ice age glacial cycles had decreased amplitude resulting in a less pronounced dichotomy of climates during glacial and interglacial periods. As such sea level changes associated with the glacial cycle were less pronounced pre-900 Kya with sea levels being approximately 70m lower than the present day, whilst after 900 Kya sea levels dropped at least 120m during glacial periods when compared to the present day. Dispersal across the tropics in the Pacific

Ocean in temperate fishes is primarily associated with cooler glacial periods allowing dispersal across the barrier (Burridge, 2002), suggesting that dispersal across the Tropical Barrier in *S. salpa* could also be associated with historical glacial periods. However, lower sea levels during the glacial period 900 Kya (and subsequent glacial maxima) would have resulted in the loss of a large area of habitat given the thin continental shelf in the region (particularly the Gulf of Guinea); resulting in substantial loss of habitat for *S. salpa* given its limited depth range of 70m (Bauchot and Hureau 1990). Therefore the change from 41 to 100 Kyr glacial cycles potentially decreased trans-tropical barrier migration possibilities during the cooler glacial periods, resulting in the observed earlier date of divergence in *S. salpa* (than for *L. mormyrus*).

For *L. mormyrus* the phylogenetic analyses suggest that the closest ancestor for Clade III to be Clade I, a clade found in both Angolan and South African waters. Such an inference is also supported by the dates for time since divergence, the two southern Atlantic *L. mormyrus* clades (Clades I and II) diverged earlier (approximately 600- 800 Kya; Tables 3.8 and 3.21), whilst Clade III (North East Atlantic) diverged from Clade I more recently at 378 Kya. Given the lack of haplotype sharing between Clade I individuals in Angola and South Africa any secondary contact between the two regions appears (comparatively) more recent than divergence between Clades I and III (see below). As such this suggests a scenario of colonisation of *L. mormyrus* directly from South Africa to the North East Atlantic. However confirmation of such a relationship could not be confirmed by IMa2 analysis since convergence across chains could not be obtained.

The estimated time of divergence between the North East Atlantic Clade III and Clade I (378 Kya) points towards the end of the Hoxnian interglacial period. The Hoxnian period was characterised by high sea levels (10m higher than present) and considerable leakage (greater than the present day) from the Agulhas Current (in South Africa) into the Atlantic (Peeters *et al*., 2004). Agulhas leakage is the transfer of warm salty Agulhas Current water into the southern Atlantic. This leakage is in the form of eddies known as Agulhas Rings which peel off the Agulhas Current as it retroflects off Cape Agulhas (Lutjeharms, 2006). Present day Agulhas rings are faunally comprised of warm temperate and tropical foraminifera typical of the southwest Indian Ocean (Peeters *et al*., 2004) and has been suggested as transporting larval Anchovy (*Engraulis capensis*) from the Benguela Current, resulting in a notable decrease in catch (Duncombe Rae *et*

al., 1992). From a phylogeographic perspective colonisation of the Atlantic (by Indian Ocean fishes) via Agulhas leakage has been suggested for Gobies (Rocha *et al*., 2005), Pygmy Angelfishes (Bowen *et al*., 2006) and the scalloped hammerhead shark (Duncan *et al*., 2006). Briggs and Bowen (2013) identify these colonisation events as being 'pulses' which correspond to interglacial periods. Agulhas leakage can feed either directly into the Benguela Current or alternatively move northwest into the warm south equatorial current (Gordon, 2003). This suggests four routes for colonising the Atlantic Ocean via Agulhas leakage.

1. Colonising Angola from Agulhas leakage feeding directly into the Benguela Current.
2. Colonising the coast of Brazil via the South Equatorial current.
3. Colonising the North East Atlantic directly through Agulhas leakage which feeds directly into the South Equatorial Current and then feeding into the Equatorial Counter Current (colonising the North African coast).
4. Colonising the mid-Atlantic islands and using them as stepping stones to colonise the North East Atlantic.

Rocha *et al*. (2005) identify the mid-Atlantic islands stepping stone route as being the most likely scenario for tropical reef associated Gobies (genus *Gnatholepis*) during the previous Ipswichian interglacial 145 Kya (Rocha *et al*., 2005). *L. mormyrus* is not found in any of the central Atlantic islands occurring only in the Canary and Cape Verde Islands in the northern Atlantic, ruling out a stepping stone scenario (although historical populations of *L. mormyrus* in the Mid Atlantic islands which have since become extinct cannot be ruled out). Instead the most likely scenario of colonising the North East Atlantic would be via the third direct approach via the South Equatorial Current and the Equatorial Counter Current. Such colonisation would have most likely been infrequent in *L. mormyrus* and Clade III being founded by only a few individuals leading to a smaller founding population in the North East Atlantic. Such a scenario would suggest a higher level of genetic drift, which could contribute to the observed divergence between Clades I and III.

During the end of the Hoxnian the Agulhas leakage reduced abruptly leading sea surface temperatures to drop quickly to 12°C in the Benguela Current region (Petrick *et al*., 2015), probably ending this mechanism of connectivity. Throughout glacial periods the Agulhas leakage was drastically reduced (Peeters *et al*., 2004; Petrick *et al*., 2015).

During the same period available habitat in the North East Atlantic for *L. mormyrus* was likely reduced southwards leading to what was essentially a glacial refugium population in Angola and (most likely) Mauritania.

Phylogeographic patterns suggest that for around 350 Kya the Tropical Barrier has been an impermeable barrier to gene flow for both *L. mormyrus* and *S. salpa*. The tropical barrier may obstruct dispersal in much the same way as suggested for *Spondyliosoma*, through a warm water barrier presented by the dominant equatorial currents in the region: the South Atlantic Equatorial Current, Guinea Current and Angolan Current (Peterson and Stramma 1991; Sherman and Hempel 2008; Verissimo *et al*., 2010). The deflection of these currents into the mid-Atlantic may also transport larvae into unsuitable habitat (Open Ocean). There are also seasonal upwelling cells off the coast of Guinea which may act as physical barriers to dispersal in larvae and adults (Bakun, 1978). Finally the narrow continental shelf in the region, coupled with the lower sea levels during glacial periods following the 900 Kya glaciation would have greatly limited the shallow water habitat available for both species, as suggested in other marine systems (Ovenden *et al*., 2009; Karl *et al*., 2012).

3.4.4 Present day genetic structuring across the Benguela Current

Both *L. mormyrus* and *S. salpa* exhibit significant genetic divergence between South African and Angolan populations, and is suggested to have resulted in the observed clades I and II respectively. The IMa2 results (Table 3.21 and Figure 3.22) suggest that this divergence was subsequently followed by asymmetrical secondary contact north from South Africa into Angola 455 Kya for *L. mormyrus* and 262 Kya for *S. salpa*. This pattern is reflected in population structure tests in both mitochondrial and microsatellite data sets (*L. mormyrus* Φ_{ST} CR = 0.092, COI = 0.220, and microsatellite F_{ST} = 0.025-0.039; *S. salpa* Φ_{ST} CR= 0.444 and COI= 0.304, and microsatellite F_{ST} = 0.028-0.031), and Bayesian clustering analysis of nuclear data for *L. mormyrus*. Other fish species studied across the Benguela region exhibit higher levels of mtDNA population structuring than observed for *S. salpa* and *L. mormyrus*: F_{ST} = 0.832-0.950 in *Diplodus capensis* (Henriques 2012); F_{ST}= 0.78 in *Lichia amia* (Henriques *et al.,* 2012); F_{ST} = 0.89 in *Atractoscion aequidens* (Henriques *et al.,* 2014). These differences in levels of differentiation may indicate longer periods of isolation of the two populations or more historically recent gene flow in *S. salpa* and *L. mormyrus* (as suggested by IMa2 analyses), or less extreme genetic drift effects (due to differences in population sizes) in

these latter species. Or the higher levels of genetic variation observed in *S. salpa* and *L. mormyrus* could have biased down the associated mtDNA F_{ST} and Φ_{ST} values (Jakobsson *et al*., 2013). However, the genetic patterns for both *S. salpa* and *L. mormyrus*, with substantial and significant differences between Angola and South Africa but no significant differences within these regions (indicated by both mtDNA and nuclear microsatellites), are compatible with gene flow restrictions associated with the Benguela Current region.

Both species yielded similar microsatellite F_{ST} values among populations across the Benguela Current (Tables 3.11 and 3.20). Slightly higher microsatellite F_{ST} values were reported for *Diplodus capensis* (0.039-0.048; Henriques, 2012), and higher values for congeneric taxa separated by the Benguela Current for up to 2 Ma, such as *A. aequidens* (F_{ST} = 0.050-0.060; Henriques *et al*., 2014) and *Spondyliosoma* (F_{ST}- 0.094- 0.103; this study). However, in contrast to *L. mormyrus*, *S. salpa* reported a non-significant F_{ST} comparison between the two regions (Table 3.20) and clustering analyses in STRUCTURE and DAPC failed to resolve any structuring across the Benguela Current (Figures 3.20 and 3.21). The relative nuclear genetic homogeneity observed in *S. salpa* could be due to a lack of statistical power, however this would seem unlikely given the POWSIM analyses identified the *S. salpa* data set could successfully detect a F_{ST} value as low as 0.0075 in 98.4% of tests. A technical concern when comparing mtDNA and microsatellite markers is that they could suffer differing levels of homoplasy. Size homoplasy is predicted to have more effect on allelic distributions between populations for microsatellites (due to their higher mutation rates) than mtDNA. However a simulation study by Estoup *et al*. (2002) suggests that size homoplasy will have much less effect on estimates of population differentiation (i.e. F_{ST}) than migration or genetic drift, making this an unlikely cause for the observed microsatellite structure in *S. salpa*. Sex-biased dispersal is often considered as an explanation for inconsistencies between nuclear and mtDNA data sets. This seems implausible for *S. salpa*, which is a protandric sequential hermaphrodite with most females reproducing in previous years as males. Furthermore there is no evidence of sex-biased dispersal reported for *S. salpa* throughout its range; as such sex-biased dispersal seems to be a highly unlikely explanation. For both species the microsatellite loci revealed numerous heterozygote deficits, which are likely to be due to null alleles (given their development for other species), which would violate the assumptions of HWE. Null alleles have been shown to

compromise statistical power (Chapuis and Estoup 2007). However DAPC, which does not require HWE still only identified a single cluster for *S. salpa* (Figure 3.21), suggesting that null alleles alone may not account for the lack of structuring observed in the *S. salpa* microsatellite data. Due to the 4X lower effective population size of mtDNA, the nuclear genome is predicted to approach equilibrium at a slower rate (Slatkin, 1994), making mtDNA more susceptible to (and potentially more sensitive to) genetic drift effects. As such microsatellite markers may retain signatures of historical gene flow longer than mtDNA. IMa2 analyses identified a more recent time of secondary contact in Angola for *S. salpa* (262 Kya) than *L. mormyrus* (455 Kya). This seems to indicate the nuclear homogeneity observed for *S. salpa* could be due to a lack of migration-drift equilibrium, wherein signatures of ancestral gene flow (which is more recent in *S. salpa*) mask current gene flow restrictions. Whilst *L. mormyrus* displayed a higher level of genetic structuring across the Benguela Current than *S. salpa*, the DAPC run using *find.clusters* (i.e. without priors) also failed to identify any clustering across the region, which again could suggest that the nuclear microsatellites are slower to reflect historical population subdivision. This could also partly explain why *L. mormyrus* displays a higher level of nuclear microsatellite genetic structuring across the Benguela, since IMa2 implies an older date of secondary contact allowing for more time to approach migration-equilibrium. IMa2 also identifies a low level of migration across the Benguela Current (which is slightly higher for *S. salpa*, Table 3.22), which although low may be sufficient to mask population structuring for some analyses and further hinder approach to migration equilibrium (Hauser and Carvalho, 2008).

The genetic population structuring between Angolan and South African *L. mormyrus* and *S. salpa* is most likely driven by the major biogeographic barrier in the region: the cold Benguela Current. The Benguela Current is likely to impose a barrier (both historically and in the present day) in much the same way that it does for other species in this thesis and studies elsewhere, whereby it represents a physical barrier (i.e. upwelling cells - Lessios *et al*., 2003; Lett *et al*., 2007), and / or a physiological barrier to adults or larvae as both species inhabit warm temperate coastal regions, and / or an ecological barrier whereby preferred food and habitat are not found in the region as they are replaced by a highly productive cold water upwelling ecosystem (Lett *et al*., 2007; Hanel and Tsingenopoulos, 2011). Both *S. salpa* and *L. mormyrus* as adults are relatively large fish indicating that currents associated with the upwelling cells found in

the Benguela Current region are unlikely to present a dispersal barrier. However there is little evidence of migratory activity in *L. mormyrus* throughout is range, in contrast to *S. salpa.* Adult *S. salpa* in South African waters are known to migrate eastwards along the coast to KwaZulu-Natal to spawn (Van der Walt and Mann, 1998). If most or all spawning of South African *S. salpa* occur in KwaZulu-Natal this would also limit the potential of the larvae even reaching the Benguela Current region during the pelagic phase. Additionally *S. salpa* eggs and / or larval survival may be compromised in cold water found in the Benguela upwelling system. Most dispersal in *L. mormyrus* is likely achieved at the pelagic larval stage (presumably contributing to the panmixia found within regions by the microsatellite and mitochondrial markers). As such the upwelling cells could have an effect of acting as a dispersal barrier with *L. mormyrus* and *S. salpa* larvae being swept into unsuitable habitat in the open ocean (Lessios *et al*., 2003; Lett *et al*., 2007)

3.4.5 Historical divergence and secondary contact across the Benguela Current

Among the southern African samples both species revealed two distinct clades: Clade I which occurs in both South Africa and Angola, and Clade II which occurs almost exclusively in Angola. Such a pattern indicates a past isolation of the two populations leading to divergence of the two haplotype clades. The initial dates of divergence for these clades vary between the two methods with IMa2 identifying more recent divergence times (*L. mormyrus* 588 Kya; *S. salpa* 491 Kya) than the p-distance method employed initially to date divergence (*L. mormyrus*. - 786 Kya; *S. salpa* 708 Kya). The margins of error for both methodologies are large and overlap (Tables 3.8, 3.17 and 3.21), with both estimates identifying a more recent date of divergence between clades I and II for *S. salpa*. These estimates of divergence broadly coincide with the end of the Mid-Pleistocene Transition (MPT). During the MPT the Benguela Current was comparably weaker and warmer (Marlow *et al*., 2000; Clark *et al*., 2006). However at the end of the MPT there was a strong cooling of the Benguela Current and an overall intensification of currents and upwelling cells, which have remained in place during the subsequent glacial and interglacial periods to the present day (Marlow, 2000). As such it seems most likely that dispersal across the Benguela Current after the MPT was restricted by these oceanographic features driving the observed divergence of clades I and II in both species.

For some fish species in the region the Benguela Current barrier has not been permeable, and remained a barrier to gene flow since the present oceanography of the Benguela Current ensued 2 Ma (Marlow *et al*., 2000; Krammer *et al*., 2006). For example, Henriques (2012) identified a divergence time of ~2 Ma between the two sister species *Argyrosomus japonicus* (South Africa) and *Argyrosomus coronus* (Angola), and similarly Henriques *et al.* (2014) identified a cryptic speciation event between Geelbek, *Atractoscion aequidens*, populations in Angola and South Africa, again with a divergence date of approximately 2 Ma. Collectively this suggests that permeability of the Benguela Current is dependable on species life history (e.g. dispersal potential) and sensitivity to features of the Benguela Current which restrict dispersal in coastal fauna, and so result in different patterns of divergence and evolutionary trajectories among species across the region. For example, *Diplodus capensis* and *Lichia amia* show more recent times of population divergence across the Benguela region (367 Kya (Henriques, 2012) and 222 Kya (Henriques *et al*., 2012) respectively), as does *Octopus vulgaris* (231 Kya- 1 Ma de Beer, 2014) than the clade divergence in *L. mormyrus* and *S. salpa*.

In both *L. mormyrus* and *S. salpa* Clade II is largely restricted to Angolan waters while Clade I is found in both Angolan and South African samples. IMa2 identified a pattern of historical vicariance of the ancestral southern African population between Angola (Clade II) and South Africa (Clade I) outlined above, subsequently followed by secondary contact by asymmetric gene flow spreading Clade I from South Africa to Angola whilst largely isolating Clade II in Angola. The IMa2 analysis indicates historical permeability of the Benguela Current system for both *L. mormyrus* and *S. salpa*. Specifically, secondary contact is dated to have occurred at 455 Kya for *L. mormyrus* and 262 Kya for *S. salpa*. This was then followed by a low level of migration until the present (Table 3.21). Sala- Bozano *et al*., (2009) previously identified such instances of secondary contact between highly divergent *L. mormyrus* clades seemingly separated for several million years across the Mediterranean / Atlantic divide and in the eastern KwaZulu-Natal region of South Africa. Bayesian assignment methods and F_{ST} analyses for both mitochondrial and microsatellite markers consistently clustered all Angolan individuals for both species as a cohesive panmictic population, showing that both clades now freely interbreed; as found in both the other sites of secondary contact in *L. mormyrus* (Sala-Bozano *et al*., 2009).

Investigations into the demographic history of Clade I individuals identified in Angola revealed independent demographic history in *S. salpa*. Angolan *S. salpa* Clade I individuals exhibit a younger date of expansion, approximately 49 Kya, than their South African counterparts which have a similar date of expansion as estimated for Clade I as a whole at 80 Kya (based on ARLEQUIN estimates). Both DnaSP and SITES supported this pattern with Clade I Angolan individuals having a more recent expansion time (Table 3.18). *L. mormyrus* also has an independent demographic history for Clade I in Angola which had a Texp of 176 Kya, whilst mismatch distribution analyses ruled out a scenario of population expansion for Clade I South African individuals. Collectively these results indicate the independent demographic history of Angolan Clade I individuals implied by IMa2 analyses.

Both estimates of the date for secondary divergence of Clade II and subsequent secondary contact both have large 95% CI (Table 3.21), making it difficult to pinpoint exact historical processes. Secondary contact is most likely to have been facilitated by Agulhas leakage during past interglacial periods. Agulhas leakage was high during interglacial periods, although Agulhas leakage in the present interglacial has not reached such historical levels (Peeters *et al*., 2004). The most likely route for larval migrants from Clade I into Angola would be via scenario 1- Agulhas leakage feeding directly into the Benguela Current and being carried northwards into Angolan waters. Alternatively migration may have occurred during environmental anomalies in the Benguela Current during which the barriers (e.g. upwelling cells) preventing gene flow are weakened, although support for such scenarios is lacking in the palaeoceanographic research (e.g. Petrick *et al*., 2015). Finally for *L. mormyrus* a single Clade II individual was identified in South Africa. Given the high level of interclade divergence homoplasy is unlikely to explain the rare occurrence of *L. mormyrus* Clade II individuals in South Africa. The IMa2 analysis identified a low level of migration from Angola to South Africa in *L. mormyrus* (Table 3.21); as such this seems to be the most feasible explanation for the rare Clade II individuals identified by the present study and Sala-Bozano *et al*., (2009) in South African waters.

3.4.6 Divergent Indian Ocean *L. mormyrus* Clade V

The CR haplotype network based on data from this study and Sala-Bozano *et al*. (2009) identified Clade V as being most closely related to Clade III (Figure 3.4). The absence of Clade V in the Agulhas Bank locality in the Western Cape, South Africa, indicates a

possible phylogeographic break between Agulhas Bank and Durban where the two clades coexist. The level of genetic divergence between Clade V and Clade III is approximately 24bp. This level of phylogenetic divergence has not been reported in this region for other species. Some taxa exhibit population structuring between the temperate Western Cape and subtropical Kwazulu-Natal but by only a few base pairs (Teske *et al*., 2011). There are three possible explanations of why this 'break' occurs here:

Sample error/ hybridisation. Samples could possibly be erroneously sampled, e.g. one of the other *Lithognathus spp.* which occur in South Africa (*L. lithognathus* and *L. aureti*). However given the high divergence at least for COI between the species in the genus (approximately 9%), that the two divergent *L. mormyrus* clades freely interbreed (Sala-Bozano *et al*., 2009) and that the species occur in sympatry make this the most unlikely of the three scenarios. Likewise a recent revision of the Sparidae family phylogeny questions whether *L. mormyrus, L. lithognathus* and *L. aureti* are even in the same genus (Santini *et al*., 2014).

Gradient from temperate to tropical waters. The apparent 'break' occurs across a well-known gradient from the temperate Western Cape to the tropical Kwazulu-Natal (Teske *et al.,* 2011). Therefore it is feasible populations may become adapted to warm temperate waters (i.e. Western Cape) and tropical waters (i.e. KwaZulu-Natal) as such Clade I would be adapted to temperate waters (and thus restricted to Western Cape) and Clade V adapted to tropical waters (and also respectively restricted). However such a significant phylogeographic break within a fish species has not been reported in the region before, but has been reported in other taxa: the coastal snail *Nassarius kraussianus* exhibits a clade adapted to the cooler Western Cape and a second clade adapted to warmer waters in the Eastern Cape (Henshilwood *et al.*, 2004). This pattern however does not hold in *L. mormyrus* as Clade I individuals found in the Agulhas locality are also found in Durban freely interbreeding (Sala-Bozano *et al*., 2009), thereby opposing the adaptability hypothesis.

Recent incursion. The clade could have migrated recently from elsewhere in the species range in the Indian Ocean, e.g. Madagascar, Oman or India. As a recent incursion clade V individuals have not penetrated as far west as the Agulhas Bank. The finding mirrors the discovery of a divergent sympatric lineage of *Octopus vulgaris* from the Durban area (Teske *et al*., 2007). The occurrence of divergent clades in Durban, one of the

busiest ports in Africa, brings about the possibility that Clade V individuals may have been introduced anthropogenically through ship ballast waters. This scenario may seem reasonable for octopus juveniles or eggs but would seem unlikely for a benthopelagic coastal fish such as *L. mormyrus*. Sala-Bozano *et al.* (2009) find it more probable that the divergent *L. mormyrus* lineage in Durban arrived by irregular inputs of oceanic waters from Eastern Madagascar (Gopal, 2006).

3.5 Conclusions

There has been increased focus on the need to delimit biogeographic units that allow for better conservation strategies in both the terrestrial and marine realms (Olsen *et al*,. 2001; Wilkinson *et al*., 2009). Here, both *L. mormyrus* and *S. salpa* report genetically divergent biogeographic units that have been shaped by historical climate change and coincide with recurrent oceanographic / habitat barriers. Such units represent interesting models for future studies aimed at understanding evolutionary divergence (e.g. Roux *et a*l., 2016). The location and permeability of biogeographic boundaries are also important to predictions of how the distribution of, and interactions among, species might change in response to climate change (Potts *et al*., 2015; Parmesan *et al*., 2005). A striking feature of the results for both species here was the historical asymmetrical gene flow from South Africa to Angola. In addition to showing, for the first time, historical permeability of the Benguela Current Barrier, this also indicates that responses to a warming ocean may be different than what is usually expected. Specifically, instead of species gradually shifting towards the poles (e.g. Sunday *et al*., 2012) this suggests that in the Benguela Current region there may be jumps across retention zones. As species distributions tend to be pinned to oceanographic boundaries, the results highlight how the Benguela Current Barrier should be a key region of study for species range shifts.

Chapter 4- Genetic divergence of two congeneric Sparids across the Benguela Current: *Diplodus cervinus* and *Diplodus hottentotus*.

4.1 Introduction

Within the family Sparidae there are 35 genera and 118 species described (Hanel and Tsigenopoulos, 2011). The genus *Diplodus* comprises 12 species, for which many sub-species have been described based on geographical differences and (often subtle) morphological variation (Hanel and Tsigenopoulos, 2011). While there is a general consensus relating to the taxonomy of the genus, Heemstra and Heemstra (2004) have suggested that many sub-species described around the Benguela system should be raised to full species status. Such a reclassification has been supported by genetic data described for *Diplodus capensis* (formerly *Diplodus sargus capensis*), for which mtDNA identified reciprocal monophyly with a time of most recent common ancestor (tmrca) of 1.8 Ma from the North East Atlantic congeneric *D. sargus* (formerly *D. s. sargus*; Henriques, 2012). Interestingly, the taxonomic break between the two congeners was identified as being between Angola and the North East Atlantic, with a comparatively more recent divergence across the Benguela Current barrier (tmrca = 367 Kya; Henriques, 2012). Correct taxonomy is fundamental to informed conservation; similar genetic based investigations are thus needed for other *Diplodus* taxa in the region. Demographic-genetic information is also directly relevant to fisheries sustainability.

The *Diplodus cervinus* complex was previously described as comprising three subspecies: *Diplodus cervinus cervinus, D. c. hottentotus* and *D. c. omanensis* (Indian Ocean endemic, Oman), but these taxa are now regarded as separate species (*D. cervinus, D. hottentotus* (Heemstra and Heemstra, 2004) and *D. omanensis* (Bauchot and Bianchi, 1984; Amir *et al*., 2013)). Present known species distributions are outlined in Figure 4.1. *Diplodus hottentotus* is the only species with a distinct break in its distribution, around the Benguela Current, where the Angolan and South African

populations are presumed to both correspond to *D. hottentotus* (Bauchot and Bianchi, 1984; Heemstra and Heemstra, 2004). With no records of this species along the Namibian or South African west coast, it has been suggested that the southern Angolan and South African populations of *D. hottentotus* may be isolated by the cold water marine biogeographic barrier formed by the Benguela Current (Floeter *et al*., 2008). Such population isolation in *D. hottentotus* might be expected as high levels of genetic divergence have been described among fish populations (e.g. *D. capensis* - Henriques, 2012; *Lichia amia* – Henriques *et al.*,2012; *Argyrosomus inodorus* - Potts *et al.,* 2014) and subspecies (*Atractoscion aequidens* - Henriques *et al.*,2014) across the Benguela Current system. The Benguela Current has also been implicated as a driver of speciation in fish populations across the region (e.g. *Argyrosomus japonicus* and *Argyrosomus coronus*- Henriques, 2012).

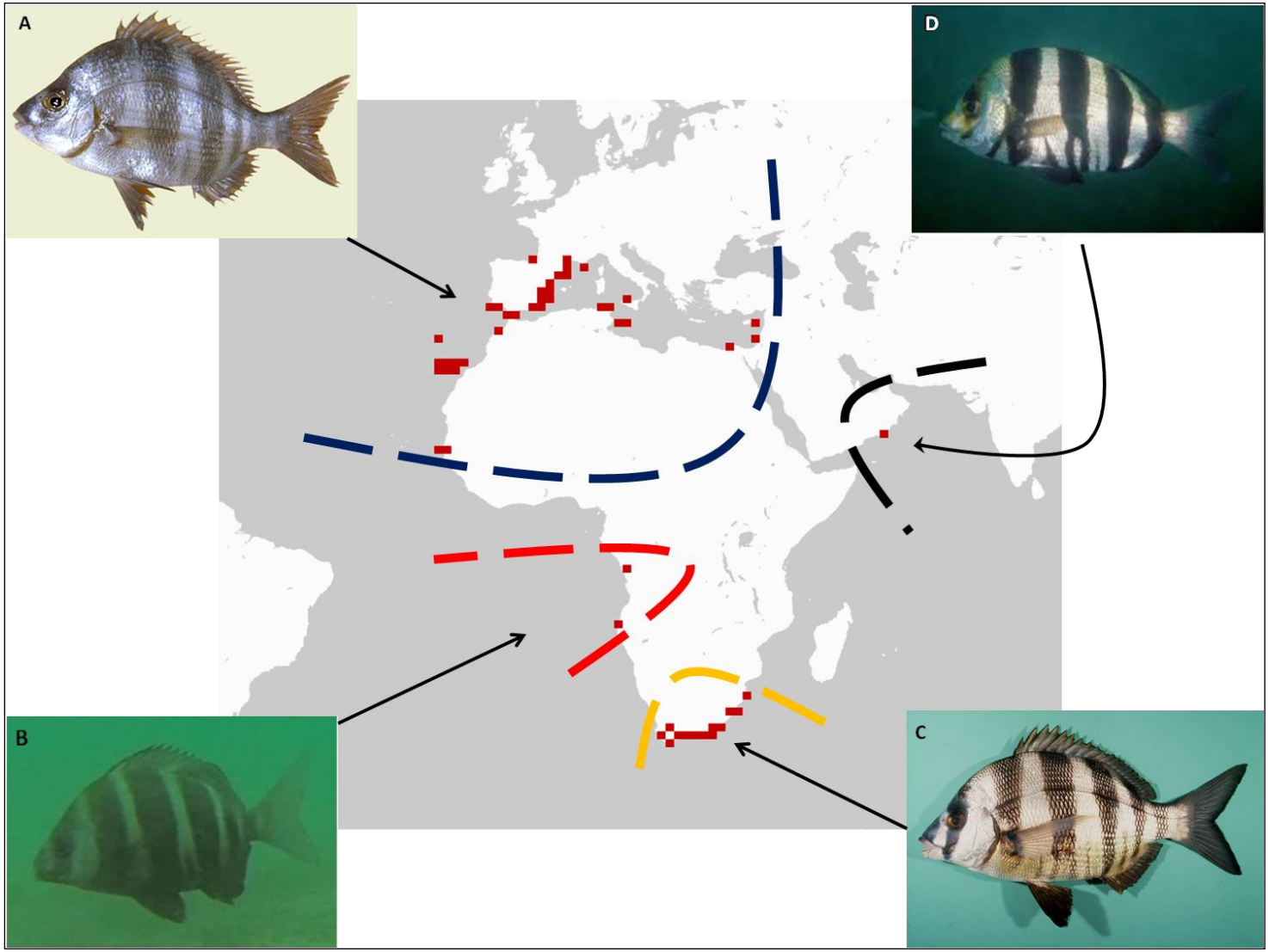

Figure 4.1 Geographical distribution maps for the three proposed *Diplodus cervinus* species and the isolated Angolan population. A= *D. cervinus*, B= isolated Angolan population which is presumed to be *D. hottentotus*; C= *D. hottentotus*, endemic to South Africa and D = D. *omanensis* which is endemic to the Gulf of Oman. Distribution map from GBIF, available at: http://www.gbif.org/species/5712108

A number of fish groups, such as *Diplodus*, are highly diversified and taxonomically complex. As such, extensive efforts are needed to elucidate their cryptic diversity. Hebert *et al.* (2003) proposed a DNA barcoding system for animals based on COI. The successful identification of species using this approach has been shown to be high in fish species (from 80-100%) for both marine and freshwater taxa (Nwani *et al*., 2011; Pereira *et al*., 2013). However, inferences based on COI alone may be compromised by the idiosyncratic behaviour of the mtDNA genome (Brower *et al*., 1996; Dupuis *et al*., 2012; Weese *et al*., 2012; Collins and Cruickshank 2013). Collins and Cruickshank (2013) also point out that there is a fundamental difference between relatively crude single locus based 'species discovery' and multilocus / integrative methods of 'species delimitation' (Sites and Marshall, 2004). The present study employed a combination of COI sequencing, which permitted integration with data for *D. cervinus* from Europe (Turkey) and nuclear microsatellite analysis to investigate taxonomic uncertainties pertaining to the relationship between hitherto described *D. hottentotus* in Angola and South Africa. Genetic patterns are also interpreted in light of phenotypic and ecological data for both regions (Winkler, 2013) towards a holistic assessment of eco-evolutionary status.

Spatial patterns of self-recruitment and connectivity are key factors shaping the dynamics of marine populations and how they respond to natural and/or anthropogenic disturbances (Hastings and Botsford 2006). For harvested species, failure to identify independent (self-recruiting) population units can lead to local over-fishing and ultimately severe declines. Self-recruitment and connectivity also determine the efficacy of management strategies, such as marine protected areas (MPAs) that are being increasingly implemented as tools to simultaneously achieve both fisheries management and biodiversity conservation objectives (McCook *et al*., 2009; McCook *et al*., 2010). Microsatellite markers have revealed significant structuring on surprisingly small geographical scales for a number of fish species with high dispersal potential and are thus used in the present study to investigate fine scale (intra-regional) population structuring among South African samples where the species is recreationally harvested.

4.1.1 Species biology

Diplodus cervinus is a demersal marine fish forming schools of 4 or 5 individuals of different sizes (Luther and Fiedler, 1976). This species is benthopelagic in shallow shelf seas, occurring on rocky bottoms mainly from 30 to 80 m in depth, but can also

occur down to 300 m on muddy bottoms. *Diplodus cervinus* is a protogynous hermaphrodite (Pajuelo and Lorenzo 2001; Pajuelo *et al.*, 2003a; 2003b). Sexual maturity occurs at 27.3 cm (TL, approximately four years of age) and males mature at around 32.7cm (approximately 5 years of age; Pajuelo *et al.*, 2003a). Spawning in *D. cervinus* extends from spring to summer, peaking from May to June (Pajuelo *et al.*, 2003a; 2003b). During spawning season, adults form schools of three to eight individuals, exhibiting polygamy with small groups formed consisting of a dominant male and several females (Pajuelo *et al*., 2003a).

Diplodus cervinus is an important commercial species in the Canary Islands, harvested with traps between three and 70 m depth year round with landings changing from season to season (Pajuelo and Lorenzo, 2001; Pajuelo ., 2003a, 2003b). However, overfishing has resulted in changes of abundance with a reduction of 85% to the unexploited equilibrium level. In 2003 the length of first capture was less than the length at maturity, with 58% of the catch being smaller than mature length. Although *D. cervinus* endures less pressure than other fish from fisheries its tendency to be resident, having a relatively late age of maturity and being a protogynous hermaphrodite could make it susceptible to overfishing throughout its range.

Diplodus hottentotus is a resident species inhabiting rocky habitats from the surf-zone to canyons at depths of 120 m (Mann, 1992; Heemstra and Heemstra, 2004). Juveniles are highly resident and found on shallow subtidal reefs, subtidal gullies, rock-pools and estuaries (Beckley, 1983, 1985; Bennett, 1987; Beckley and Buxton 1989; Mann 1992; Watt-Pringle, 2009). *Diplodus hottentotus* eggs and larvae are primarily found inshore (Connell, 2012). Adults feed on a range of benthic invertebrates including crustaceans and gastropods, but specialises on polychaetes and amphipods (Mann, 1992; Winkler, 2013).

Diplodus hottentotus is a rudimentary hermaphrodite (protogynous) with a length at 50% sexual maturity of 24 cm FL and an age at 50% maturity of approximately eight years (Winkler *et al.,* 2014). Spawning takes place from August to December, peaking in October in the Eastern Cape (Mann and Buxton, 1998) and occurring on inshore reefs recorded off the KwaZulu-Natal south coast (Connell, 2012) and in False Bay (Brownell, 1979). The species has a maximum age of 33 years (Mann and Buxton, 1997), a maximum recorded length of 60 cm (Heemstra and Heemstra, 2004) and maximum recorded weight of 5.4 kg (SADSAA, 2012).

4.2 Methods

4.2.1 Sampling and DNA extraction

168 individuals of *Diplodus hottentotus* were collected from eleven sampling sites in Angola and South Africa, plus two outgroup individuals of *Diplodus cervinus* from Turkey (see Table 4.1 and Figure 4.2). Samples were obtained from a mixture of recreational angling and local fish markets. A fin clip was removed from each individual and preserved in 95% ethanol. Total genomic DNA was extracted following the phenol-chloroform method described by Sambrook *et al*. (1989) and visualised on a 1% agarose gel.

Table 4.1 Sampling strategy for *D. cervinus spp.* where Sample size is the number of individual fish sampled in the present study. COI is the total number of individuals sequenced for COI per site and Microsatellites is the total number of individuals genotyped per site. * denotes COI sequences derived from GenBank

Region	Site (ID)	Sample Size	COI	Microsatellites
Angola	Benguela (BEN)	2	2	-
	Lucira (LUC)	9	7	-
	Namibe (NAM)	1	1	-
	Flamingo (FLA)	50	23	40
South Africa	Struisbaai (STR)	-	1*	-
	Koppi Alleen (KOP)	8	5	-
	Tsitsikamma (TSI)	34	10	22
	Port Elizabeth (PEL)	38	9	38
	Port Alfred (PAL)	16	6	-
	West Kleinmonde (WEK)	7	5	-
	East London (EAL)	1	1	-
	Cwebe (CWE)	2	1	-
	Durban (DUR)	-	1*	-
	Kwazulu- Natal (KZN)	-	2*	-
Europe	Turkey (TUR)	2	2/21*	-

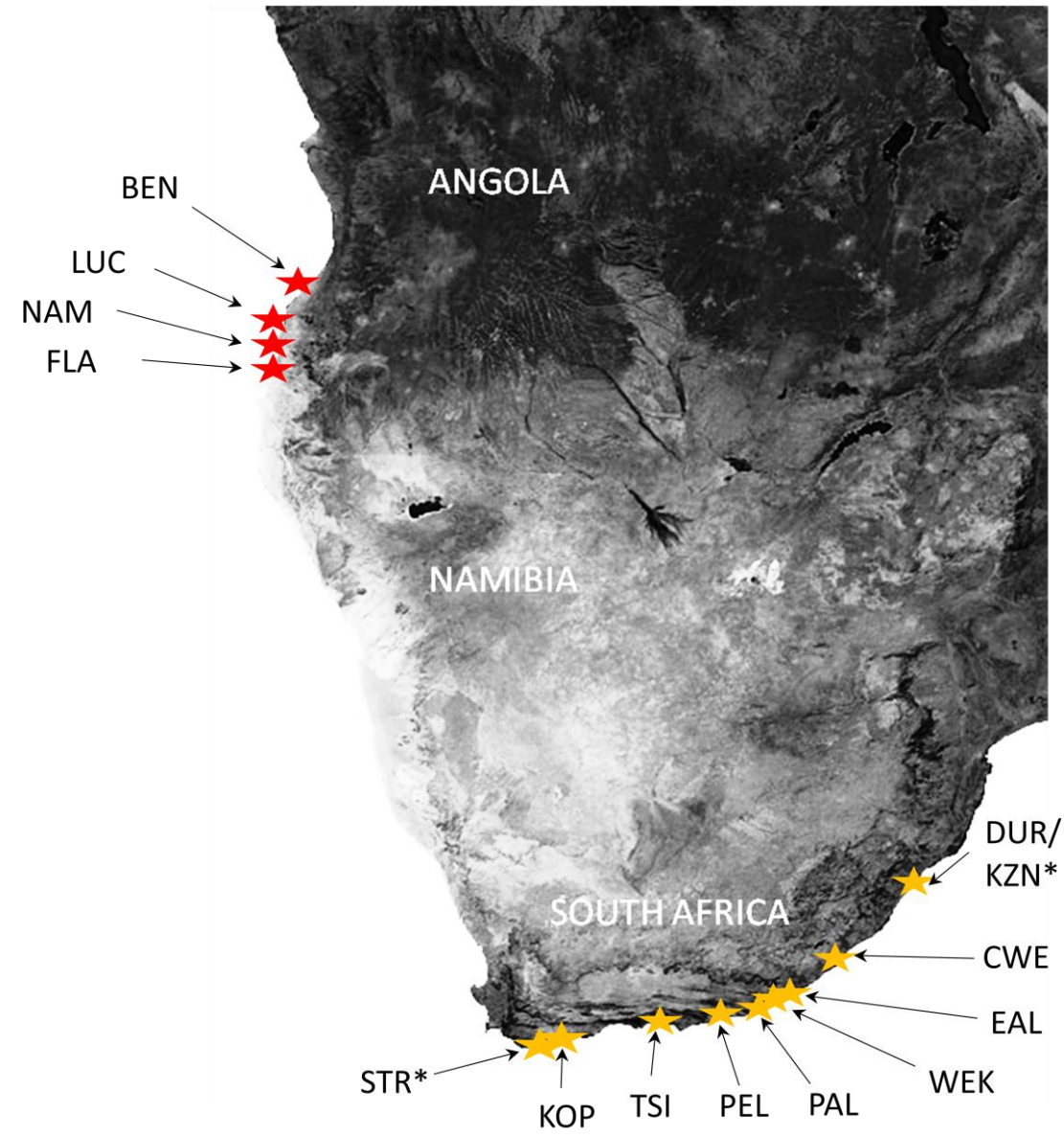

Figure 4.2 Sampling strategy for *Diplodus hottentotus.* Site ID codes are as denoted in Table 4.1. * denotes COI sequence(s) derived from GenBank sampling sites.

4.2.2 MtDNA markers and analyses

A 501bp fragment of the mtDNA Cytochrome Oxidase I (COI) gene was amplified using PCR with species-specific primers DCCOIF (5' TCATTCCGAGCCGAACTAAGC 3') and DCCOIR (5' TCCTGCAGGGTCAAAGAAAG 3'). These primers were developed, using PRIMER3 0.4.0 (Rozen and Skaletsky, 2000), from initial *D. hottentotus* sequences produced by amplification using the universal fish primers COI-WF1 and COI-WR1 (Ward *et al*., 2005). MtDNA Control Region (CR) was also initially screened however, the position of the universal CR primers was found to be placed within a large repeating unit, thereby producing PCR product of varying lengths. As such the repeat region proved to be problematic for sequencing and subsequent generation of species specific primers was not feasible. As such CR sequences could not be obtained.

PCRs comprised of 10 µl of BIOMIX (BioLine), 1.0 pMol of primer (both forward and reverse), 6 µl of template DNA and 2 µl of sterile distilled water giving a total reaction volume of 20µl. All PCRs were performed using a C1000 Thermal Cycler (Bio-Rad) using the following reaction conditions: 120 s at 95°C, then 40 cycles of 30 s at 94°C, 30 s at 50°C, 60 s at 72°C, with a final extension step of 120 s at 72°C. Resultant products were then verified on a 2% agarose gel, purified using SureClean (BioLine) and sequenced using an ABI 3730 DNA analyser (Applied Biosystems ®).

COI sequence chromatograms were prepared for analyses as per Chapter 2 methodology. Genetic diversity indices and genetic structure tests (and visualisation) were performed as per Chapter 2. mtDNA COI Power Analysis was performed in POWSIM with methods as described in Chapter 2.

To test for signals of past population expansion sequence mismatch distributions and the expansion parameter τ (Rogers and Harpending, 1992), were estimated using the COI data set in the software packages ARLEQUIN, DnaSP and SITES. Analyses were performed using the same parameters as in Chapter 3 methodology.

4.2.3 Microsatellite DNA markers and analysis

Following testing of 18 published nuclear microsatellite Sparid loci a subset of six polymorphic loci (see Table 4.2) which provided consistent PCR amplification were used to assess nuclear genetic variation within two population samples from South Africa (Tsitsikamma and Port Elizabeth) and one sample from Angola (Flamingo).

Samples were amplified using PCR under the following conditions: 300s at 95°C, then 30 cycles of 30s at 92°C, 30s at a primer and species specific annealing temperature (see Table 4.2) and 30s at 72°C, and a final extension step of 72°C for 120s. All reactions used the following reaction mix: 5 µl of BIOMIX (BioLine), 0.5 pMol of primer (both forward and reverse), 3 µl of template DNA and 1 µl of sterile distilled water giving a total reaction volume of 10µl. Alleles were separated using an AB3730 DNA analyser and allele identity inferred using Peak Scanner 2.

Microsatellite genetic variation analyses were performed as per Chapter 2. Genetic structuring tests carried out on the microsatellite data set included two reliant on Hardy-Weinberg Equilibrium (F_{ST} and Bayesian clustering analysis in STRUCTURE) and also non- Hardy Weinberg equilibrium reliant DAPC carried out in *adegenet* as per chapter 2 methods. Microsatellite Power Analysis was performed in POWSIM with methods as described in Chapter 2.

Table 4.2 The seven microsatellite loci used in the present study, showing the original species the primers were developed for, repeat motif, forward and reverse primer sequences and the optimised Ta (°C) used in the present study for *D. cervinus* spp.

Microsatellite ID	Original species	Reference	Primers F	Primers R	Ta °C
DsaMS16	*Diplodus sargus*	Perez *et al*. (2008)	F: AGTCAAACCTCGGCATCAAGCGGGTA	R: ACGAGGAGCTCTGACTTCTGATTCGTT	55
DsaMS27	*Diplodus sargus*	Perez *et al*. (2008)	F: GCTCACTGTGCTGGCTCCACATCACC	R: GCGCTGTGCTTGCTGTCGGAGA	55
DsaMS34	*Diplodus sargus*	Perez *et al*. (2008)	F: AGATCAGATTTGCTGTGATAGCGTCCAAAG	R: ACTCCTGCAGCTCCTCCTGGGCTTC	55
Dvul33	*Diplodus vulgaris*	Roques *et al.* (2007a)	F: GCCGGGCTCGACATTGACACTGAA	R: GCAGCCAGCAGAGCTTAAAGAACT	50
Dvul4	*Diplodus vulgaris*	Roques *et al.* (2007a)	F: GCGGTTATGTATACGTTGCGTTTA	R: TTGGCGTTGAACAGAAGTCAGACA	55
Dvul84	*Diplodus vulgaris*	Roques *et al.* (2007a)	F: GCTCGACGTGCACTCTGCCCTTGA	R: ATTCCCCAAATCCAGCACTCACAT	55
Omel58	*Oblada melanura*	Roques *et al.* (2007b)	F: GGCATTATTGTTCCATCATTACTCC	R: ATGGCATACAACCTGCATCAGAAG	55

4.3 Results

4.3.1 MtDNA COI diversity, phylogeography and population structuring

Pruning of mtDNA sequences permitted comparison of 501 sites across 74 individuals and revealed a total of 13 haplotypes. Overall genetic diversity was h= 0.758 and π = 0.0026, being highest in Angola (h= 0.725 (SD 0.0600), π = 0.0024 (SD 0.0017)), at intermediate level in Turkey (h= 0.577 (SD 0.088), π = 0.0016 (SD 0.0013) (Table 4.3), and lowest in South Africa (h= 0.364 (SD 0.090), π = 0.0015 (SD 0.0012)) due to this region comprising one common haplotype and 4 singletons. Individual sample genetic diversities are reported in Table 4.4.

Reconstruction of the relationships amongst the resolved haplotypes (Figure 4.3) revealed a clear partitioning of haplotypes between Angola and South Africa with only one haplotype shared between both regions (Haplotype 7). Three haplotypes were identified among the Turkish samples and these were found to occupy central positions in the haplotype network with one (Haplotype 6) being the most common haplotype among South African samples, and the other two (Haplotypes 2 and 3) being the most common among the Angolan samples (Figure 4.3). A homoplasy ring is also evident in the median joining network, the dashed line represents the suggested break point utilising coalescent theory (Posada and Crandall, 2001).

For population structuring tests between samples within regions and between regions only sample sites with five or more individuals are included. The clear partitioning of haplotypes between Angola and South Africa translated to highly significant pairwise Φ_{ST} tests between samples from both regions (Φ_{ST} =0.30-0.59) and between regions pooled (Φ_{ST} = 0.50), but with no significant values between samples within regions (Tables 4.5, and 4.6.). The Turkey sample also displayed significant Φ_{ST} values against Angola and South Africa, but with much lower values against Angola ((0.111) than South Africa (0.475). AMOVA supported separation between Angolan and South African populations with between-region variance (Angola/ South Africa) explaining 41.45% (P= 0.004) of frequency variation within the model (Table 4.7).

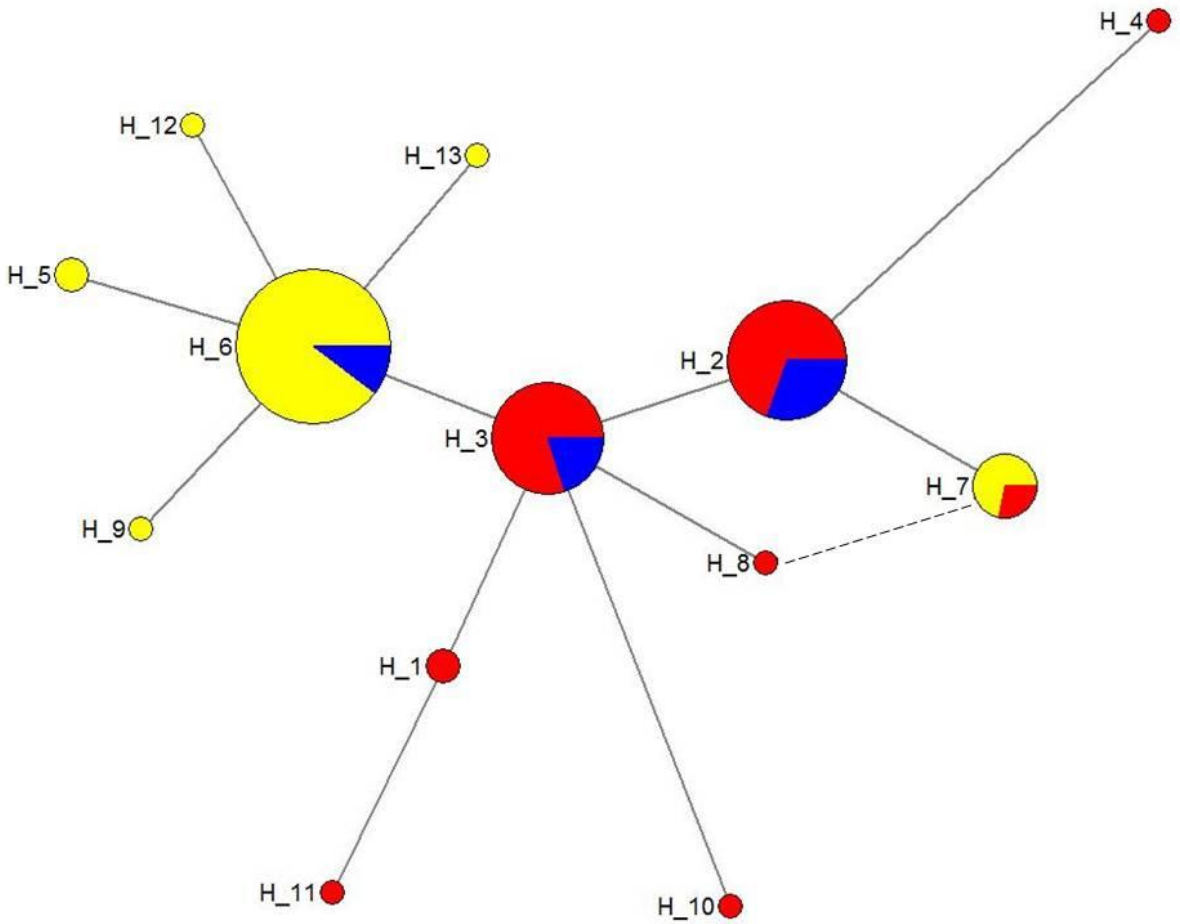

Figure 4.3 Reconstructed median-joining haplotype network for *D. cervinus spp*. based on 501bp of mtDNA COI. Node sizes are proportional to the observed number of individuals bearing that haplotype. Colouring refers to sample regions with Yellow corresponding to South Africa, Red to Angola and Blue to Mediterranean. Dashed linkage represents proposed breakage point of homoplasy ring.

Table 4.3 Genetic diversity for *Diplodus cervinus* spp. mtDNA COI sequences for the regions. n: sample size; H: haplotype number; PH private haplotype number; h: haplotype diversity; π: nucleotide diversity; τ: tau which is derived from three software packages- ARLEQUIN (including 95%CI), DnaSP and SITES; PSSD: the probability that the empirical distribution of mismatches was significantly different than the distribution simulated under a demographic expansion model (P= P-value); Raggedness: Harpending's raggedness index; Texp: Time since expansion (with 95% CI for ARLEQUIN estimates);*D*: Tajimas D and F_s=Fu's FS. Bold indicates statistically significant values. Note all regional sample sites are used including GenBank COI Sequences.

	Angola	**South Africa**	**Turkey**	**Total**
N	34	40	23	97
H	8	6	3	13
PH	5	4	0	-
h (SD)	0.725 (0.060)	0.364 (0.090)	0.577 (0.088)	0.758
π (SD)	0.0024 (0.0017)	0.0015 (0.0012)	0.0016 (0.0013)	0.0026
τ (Arleq; 95% CI)	1.230 (1.903; 0.406)	3.000 (4.152; 0)	1.033 (2.488; 0.000)	1.436 (1.988; 1.021)
PSSD (P)	0.022 (0.316)	0.002 (0.664)	0.002 (0.694)	0.002 (0.305)
Raggedness (P)	0.270 (0.484)	0.072 (0.376)	0.072 (0.849)	0.052 (0.410)
Texp (Arleq)	102 Kya (158; 34)	250 Kya (345; ongoing)	86 Kya (207; Ongoing)	119 Kya (165; 85)
τ (DnaSP)	0	1.186	0.798	1.329
Texp (DnaSP)	Present Day	99 Kya	66 Kya	111 Kya
τ (SITES)	0.552	0.801	0.291	0.873
Texp (SITES)	46 Kya	67 Kya	34 Kya	73 Kya
***D* (P)**	-1.185 (0.123)	**-1.479 (0.049)**	1.056 (0.832)	-1.283 (0.085)
F_S (P)	**-3.048 (0.029)**	-2.139 (0.074)	0.840 (0.677)	**-5.599 (0.015)**

Table 4.4 Genetic diversity inferred from mtDNA COI sequences. n: sample size; H: haplotype number; h: haplotype diversity and π: nucleotide diversity. ID codes are as denoted in Table 4.1. Note: sites with a limited number of individuals (< 5) are omitted.

	LUC	**FLA**	**KOP**	**TSI**	**PEL**	**PAL**	**WEK**
n	7	23	5	10	9	6	5
H	4	6	2	2	4	3	1
h (SD)	0.714 (0.181)	0.727 (0.065)	0.400 (0.237)	0.200 (0.154)	0.583 (0.183)	0.600 (0.215)	-
π (SD)	0.0027 (0.0021)	0.0023 (0.0017)	0.0024 (0.0021)	0.0004 (0.0006)	0.0022 (0.0018)	0.0027 (0.0022)	-

Table 4.5 Pairwise Φ_{ST} values based upon the mtDNA COI data, between *D. cervinus spp.* samples. Bold indicates statistically significant F_{ST} values. Site ID codes are as denoted in Table 4.1. Note sites with a limited number of individuals (< 5) are omitted.

	LUC	FLA	KOP	TSI	PEL	PAL	WEK
FLA	-0.019	-					
KOP	**0.424**	**0.362**	-				
TSI	**0.576**	**0.476**	-0.037	-			
PEL	**0.356**	**0.321**	-0.12	0.027	-		
PAL	**0.34**	**0.303**	-0.16	0.04	-0.141	-	
WEK	**0.591**	**0.484**	0	-0.084	0.034	0.063	-
TUR	**0.199**	**0.095**	**0.394**	**0.487**	**0.344**	**0.338**	**0.501**

Table 4.6 Pairwise Φ_{ST} values for regions based upon *D. cervinus* spp. mtDNA COI sequence data. Bold indicates statistically significant values.

	South Africa	Angola
Angola	**0.464**	-
Turkey	**0.475**	**0.111**

Table 4.7 Analysis of molecular variance results based on *D. cervinus spp.* COI sequence data. The AMOVA was structured using Angolan and South African sample regions as groups. Bold indicates statistically significant results.

Source of variation	Percentage of variation.	P-value
Among groups	**41.45**	**0.004**
Among populations within groups	0.21	0.417
Within populations	58.33	0.415

4.3.2 Demographic history

With the absence of population structuring within regions all subsequent analyses of regional historical demography used sample sites within Angola and South Africa pooled. Analyses reconstructing historical demography could not exclude a hypothesis of past demographic expansion in both Angolan and South African populations: significantly negative values for Tajima's *D* and Fu's *Fs* were obtained for both regions (Table 4.3). Mismatch distribution analyses also did not allow the rejection of the null hypothesis of demographic expansion (Table 4.3). Utilising the ARLEQUIN τ values and assuming a divergence rate of 1.2% Myr^{-1}estimated time since expansion in South Africa as ~250 Kya (95% CI 345-ongoing) and in Angola as~102 Kya (95% CI 158-34). Whilst the three estimates of expansion time (and τ) from Arlequin, DnaSP and SITES vary (See Table 4.3) they are consistent in estimating an older time of expansion

in South Africa and a more recent expansion in Angola. The Turkey population of *D. cervinus* displayed positive and non-significant Tajima's D and Fu's Fs values, but mismatch distribution analyses could not rule out a hypothesis of demographic expansion (Table 4.3), yielding τ = 1.033 (95% CI 0.000-2.473) which translated to an expansion time of ~86 Kya (95% CI 207-ongoing).

4.3.3 Nuclear Microsatellite DNA diversity and structuring

Information on microsatellite genetic variation for each sample/locus combination is provided in Table 4.8. There were no significant deviations from random associations of genotypes (linkage disequilibrium) detected for any pair of loci, either across all samples (data pooled) or in any single sample, indicating that all loci assort independently. All loci were variable in each sample with the total number of alleles per locus ranging from two (DsaMS27) to 28 (Dvul84) with an average of 8.43**.** Although levels of variability differed across loci, multi-locus variability indices were similar across all samples (Table 4.8). Significant deviations from HWE were found in 9 out of 21 locus / sample comparisons (Flamingo- 3 of 7 tests; Port Elizabeth - 3 of 7 tests; Tsitsikamma - 3 of 7 tests), in eight cases due to heterozygote deficits, whilst the Tsitsikamma / DsaMS34 comparison exhibited a heterozygosity excess (Table 4.8). Furthermore FreeNA indicated the presence of null alleles in 14 of the 21 sample / locus comparisons.

Bayesian clustering in STRUCTURE unanimously supported a model of *K*=2 according to both log probability (*P*=1 for *K*= 2, and zero for other models) and Evanno's delta K methods of interpretation for both runs with and without *LocPrior*. Clustering of individuals into both clusters followed a geographic pattern with one cluster containing all Flamingo (Angola) individuals while all Tsitsikamma and Port Elizabeth (South Africa) individuals assigned to the other cluster (Figure 4.4). The run(s) without *LocPrior* show a few Angolan individuals displaying some genetic similarity to South Africa and vice versa (Figure 4.4). The pattern of genetic structuring between Angolan and South African populations was also supported by pairwise F_{ST} values estimated with and without null allele correction, which were high (>0.23) and highly significant in all comparisons involving Flamingo against the two South African samples (Table 4.9).Comparisons between Tsitsikamma and Port Elizabeth yielded much lower F_{ST} (<0.04) which were non-significant (Table 4.9).

Table 4.8 Genetic diversity of seven microsatellite loci in *D. cervinus spp*. n- number of individuals genotyped; Na- number of alleles; AR allelic richness; H_E- expected heterozygosity; Ho- observed heterozygosity. pHWE- Hardy–Weinberg equilibrium probability. Site ID codes are as denoted in Table 4.1. Bold indicates statistically significant values.

		FLA	**TSI**	**PEL**
Total	**N**	40	22	39
DsaMS16	**N**	40	22	38
	N_a	6	4	4
	H_E	0.275	0.214	0.197
	H_O	0.300	0.227	0.158
	pHWE	1.000	1.000	0.336
DsaMS27	**N**	40	22	38
	N_a	2	2	2
	H_E	0.073	0.241	0.125
	H_O	0.075	0.182	0.132
	pHWE	1.000	0.324	1.000
DsaMS34	**N**	40	22	39
	N_a	5	5	6
	H_E	0.585	0.693	0.719
	H_O	0.400	0.909	0.692
	pHWE	**0.011**	**0.011**	0.766
Dvul4	**N**	32	11	37
	N_a	5	3	3
	H_E	0.393	0.255	0.105
	H_O	0.156	0.091	0.054
	pHWE	**>0.001**	**0.048**	**0.028**
Dvul33	**N**	39	22	39
	N_a	3	3	3
	H_E	0.303	0.320	0.169
	H_O	0.359	0.091	0.128
	pHWE	0.638	**0.001**	**0.029**
Dvul84	**N**	35	12	37
	N_a	18	12	17
	H_E	0.903	0.92	0.91
	H_O	0.486	0.833	0.811
	pHWE	**>0.001**	0.152	0.052
Omel58	**N**	23	7	38
	N_a	7	2	5
	H_E	0.698	0.264	0.635
	H_O	0.652	0.000	0.316
	pHWE	0.065	0.078	**>0.001**

Table 4.9 Pairwise F_{ST} values between *D. cervinus* spp. samples based on seven microsatellite loci, Below the diagonal are the uncorrected F_{ST} values calculated in FSTAT with significance assessed after permutation; above the diagonal are the F_{ST} values after correction for null alleles calculated in FreeNA, with significance being estimated by 95 % confidence after 1000 bootstrap replicates. Site ID codes are as per Table 4.1. Bold indicates statistically significant F_{ST} values.

	FLA	**TSI**	**PEL**
FLA	-	**0.284**	**0.236**
TSI	**0.276**	-	0.017
PEL	**0.238**	0.019	-

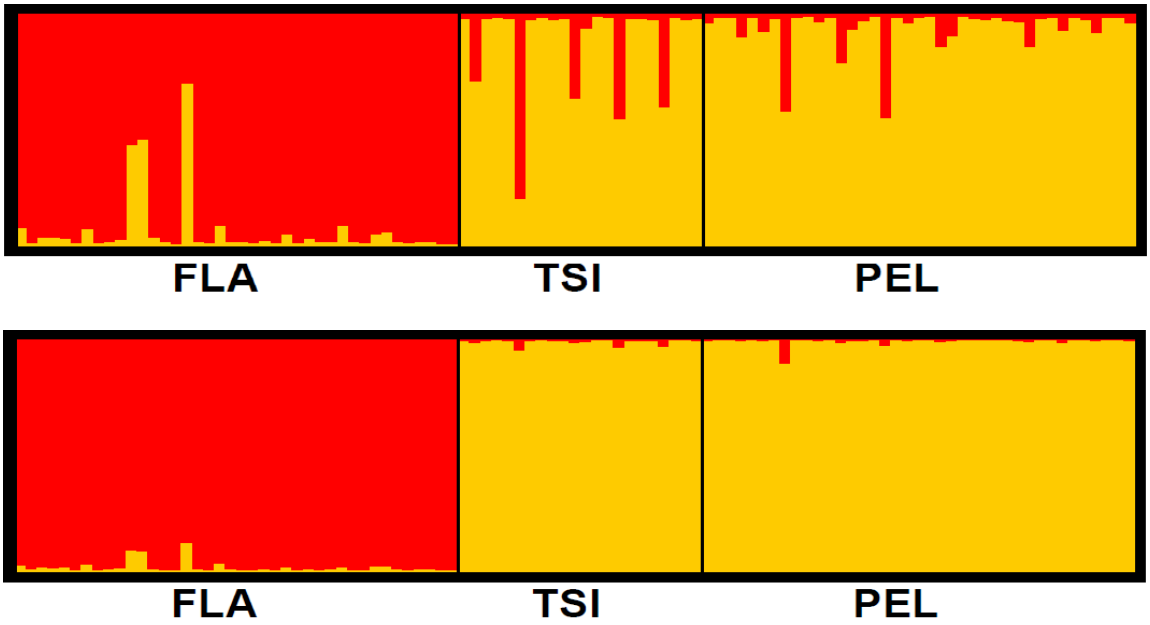

Figure 4.4 Number of genetic clusters observed within *D. cervinus spp.* populations across the Benguela region. Assignment values for each individual fish obtained from STRUCTURE, based on genotypes from seven nuclear microsatellite loci, for K = 2. TOP: Not assuming priors BOTTOM: Assuming priors. Clusters correspond geographically to Angola (Red) and South Africa (Yellow). Sampling site codes as per Table 4.1

Both the DAPC runs with and without priors identified the geographical clustering of Angolan and South Africa. For both runs the first 30 PC's were retained representing 96.5% of variability. Once plotted with the priors DAPC identified clear separation (with no overlap) of Angolan and South African individuals (Figure 4.5), with the first principle component discerning the clear difference between Angolan and South African individuals. The DAPC run without priors using the *find.clusters* function resolved a K= 2, correctly assigning all individuals geographically to either South Africa (all Tsitsikamma and Port Elizabeth individuals) or Angola (all Flamingo individuals). Once plotted the inferred clusters show separation and little overlap (Figure 4.5).

4.3.4 Power Analysis

Power analysis identified that the mtDNA COI marker presented a low Type I error probability, (Fisher P: 0.047), but had a significantly lower power to detect shallower differentiation F_{ST} = 0.010 (Fisher P: 0.145) power only reached the 95% threshold when F_{ST} = 0.125 (Fisher P: COI = 0.963). The microsatellite data again had a low Type I error (Fisher P = 0.054) and a considerably higher power than the COI markers with a high probability (Fisher P = 0.965) for detecting differentiation at F_{ST} = 0.0125.

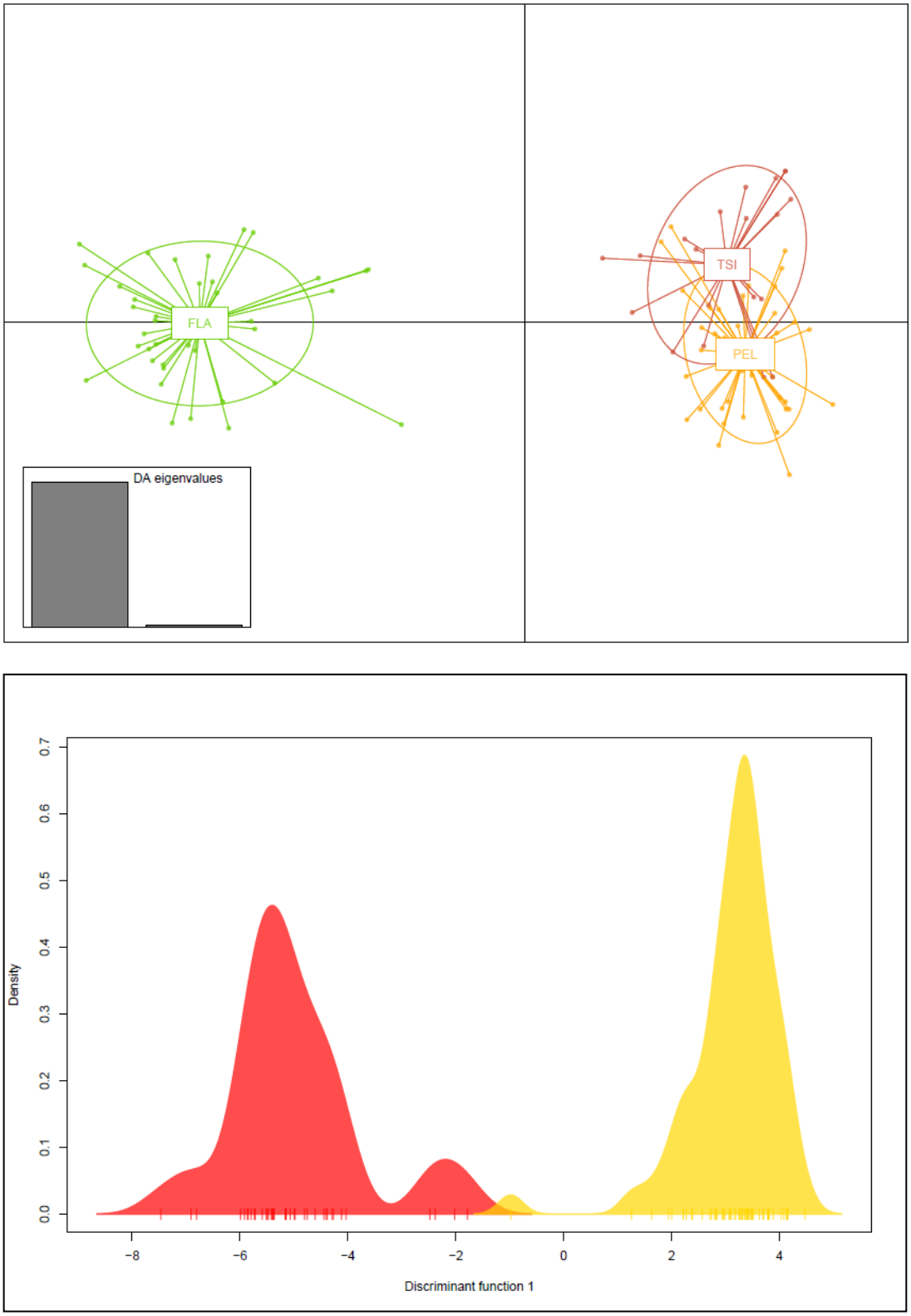

Figure 4.5 TOP: scatter plot of individuals on the two principal components of the DAPC with groups defined *a priori* as per sample site. The graph represents individuals as dots and the groups as inertia ellipses. Eigenvalues of the analysis are displayed in inset. Individual population codes can be identified in Table 4.1. BOTTOM: Graph showing results of the DAPC obtained with the *find.clusters* option (i.e. no priors) where clusters= 2, with a single discriminant function displayed on the x- axis, (with individuals represented as dashes) whilst density of individuals is plotted on the y -axis. Clusters concur geographically to Angola (Red) and South Africa (Yellow).

4.4 Discussion

The salient feature of the results is the pronounced cyto-nuclear genetic differentiation between the Angolan and South African samples. Φ_{ST} values for comparisons of samples between regions (i.e. Angola and South Africa) utilising mtDNA COI were high (Φ_{ST}= 0.496). The mtDNA haplotype network, though shallow and with only five nucleotide differences between maximally diverged haplotypes, exhibits a clear phylogeographic structure: of 13 haplotypes resolved among South African and Angolan samples only one (haplotype 7, a tip haplotype) is found in both regions. Of interest is the observation that although South African and Angolan populations share almost no mtDNA variation both regions share their common haplotypes with the Mediterranean 'outgroup' sample (Turkey). Nuclear microsatellite variation also revealed a high level of differentiation between Angolan and South African samples (F_{ST} ~ 0.22), but no differentiation among South African samples, supported by clustering analyses which provide no evidence of migrants or first generation hybrids between regions. As the analysed samples consist of individuals from multiple age / size cohorts the regional differentiation in both marker types cannot be attributed to temporal variation, for example due to sweepstakes recruitment (McKeown *et al*., 2017). The pattern and extent of genetic differentiation among regions supports a hypothesis of restricted gene flow and absence of dispersal across the Benguela Current system, as observed for other coastal fishes across this region (Henriques, 2012; Henriques *et al*., 2012; Henriques *et al*., 2014; Potts *et al*., 2014; Henriques *et al*., 2016).

4.4.1 Taxonomy

Given the existing species descriptions (*D. hottentotus* in South Africa and Angola, *D. cervinus* in North East Atlantic and Mediterranean), from a taxonomic perspective the COI dataset does not clearly resolve into the monophyletic clades expected by more traditional species concepts such as the Phylogenetic Species Concept (PSC; Cracraft, 1989). The Mediterranean population of *D. cervinus* exhibits no private COI haplotypes and all three haplotypes present in the Mediterranean are shared with, and are the common haplotypes in, the two southern African populations (two with Angola and one with South Africa (Figure 4.3)). Such a pattern undermines the premise that a defined species break between proposed *D. cervinus* and *D. hottentotus* occurs between Angola and the North East Atlantic, as recognized in the congeneric species *D. capensis* and *D.*

sargus (Henriques, 2012). As such the major phylogeographic break identified here is between Angolan and South African samples. The overall pattern of nuclear and mtDNA variation between Angolan and South African populations suggests that the sharing of haplotype 7 is due to incomplete lineage sorting rather than some form of recent introgression (see Moran and Kornfield, 1993). The occurrence of incomplete lineage sorting is also supported by the presence of haplotype 6 (a central haplotype) in both the South African and Mediterranean samples but absent from Angolan samples. Taken at face value the lack of complete monophyly between the three regions / populations would appear to indicate that *D. cervinus* and *D. hottentotus* do not represent fully diverged taxa.

However there are two prominent aspects that counter such an argument of taxonomic homogeneity across the full range of *D. cervinus* / *D. hottentotus*. Firstly, recent studies of meristic, morphometric and life history traits found significant differences between the Angolan and South African populations (Winkler, 2013). The level of morphological divergence observed between *D. hottentotus* populations is comparable to that observed between populations of *Atractoscion aequidens* across the same regions and which are suggested to be cryptic species (Henriques *et al*., 2016). Secondly, the microsatellite F_{ST} value is far greater than that observed between Angolan and South African populations of the other species in this thesis, and greater than those observed in *A. aequidens* (F_{ST} = 0.050-0.060; Henriques *et al*., 2014). Together these results indicate substantial divergence comparable to that of other species pairs identified in this biogeographically complex region.

The retention of ancestral mtDNA polymorphism between the putative *D. cervinus* and *D. hottentotus* mentioned above seems to run contrary to the observed divergences in morphology and microsatellite differentiation. Such cases of retention of ancestral polymorphism have been a major factor driving concerns as to the use of 'thresholds' or exclusivity criteria (reviewed in Sites and Marshall 2004) applied to species delimitation from genetic data (Hudson and Coyne 2002; Hudson and Turelli 2003; Moritz and Cicero, 2004, Matz and Nielsen 2005). Essentially, recently derived species will tend to go undiscovered under a reciprocal monophyly criterion since species boundaries are not faithfully reflected in a gene tree until sufficient time has elapsed post-divergence for ancestral polymorphism to be fully sorted (e.g. Hickerson *et al*., 2006). Recent studies have also shown that gene genealogies can provide information

about species divergence despite widespread incomplete lineage sorting (e.g. Degnan and Salter, 2005; Maddison and Knowles, 2006; Carstens and Knowles, 2007; Knowles and Carstens, 2007). As such there has been a move away from universal exclusivity criteria towards methods of species delimitation that might better encompass the stochastic variance of genetic processes underpinning speciation (Knowles and Carstens, 2007; Hudson and Turelli, 2003; Panchal and Beaumont, 2007). The importance of reciprocal monophyly for incipient taxa is questionable (Meier, 2008) as it may be an unrealistic scenario in many closely related groups (Funk and Omland, 2003; Zhang *et al*., 2012), and there is a recognition that coalescent depths among species will vary considerably due to differences in population size, mutation rate and time since speciation (Monaghan *et al*., 2009; Fujita *et al*., 2012). In addition to various demographic factors this higher level of lineage sorting in *D. capensis / D. sargus* compared to *D. hottentotus / D. cervinus* could be linked to the faster generation time in *D. capensis*, which matures in 1.8 years and has a maximum lifespan of 31 years (Richardson *et al.*, 2011), compared to *D. cervinus* which sexually matures later at 4.9 years and lives to a maximum age of 43 years (Mann and Buxton, 1997; Winkler, 2013).

The genetic differences among South African and Angolan samples are compatible with prolonged periods of genetic isolation and distinct evolutionary trajectories (Waples, 1998). In addition they also align readily with differences in general phenotype and morphology described by Winkler (2013). Such congruent genetic / morphological divergence has driven taxonomic reappraisals in other groups (e.g. Gobidae; Lima-filho *et al*., 2015). Regarding the use of 'hottentotus', whether for full species or subspecies status, this should be restricted to South African *Diplodus* 'cervinus' to reflect their status as distinct 'species- like units' (*sensu* Collins and Cruickshank, 2013), and the Angolan population of *Diplodus* 'cervinus' should be assigned its own designation. Such a redefinition can be made conveniently due to the clear geographical separation of both units, and their separation from the North East Atlantic / Mediterranean units.

4.4.2 Differences in genetic diversity between regions

Although levels of microsatellite variability were similar between both regions (Angola and South Africa), the South African samples exhibited markedly lower mtDNA diversity. This difference between marker types could be attributed to greater drift effects at mtDNA owing to its ¼ effective population size as well as the higher mutation

rates of microsatellite loci. A similar pattern of reduced mtDNA variation among South African samples has been reported for *Lithognathus mormyrus* and *Sarpa salpa* (see Chapter 3). Such a cross-species pattern could be due to differences between regions in historical climate-induced contraction / colonisation (e.g. founder effects) dynamics as suggested in Chapter 3. However, in other species (e.g. *Spondyliosoma* - this study; *D. capensis* - Henriques, 2012) South African samples exhibit higher levels of mtDNA variation than Angolan counterparts. It cannot be ruled out that the distinct fishing pressures in both regions have also influenced levels of genetic variation. Specifically, while South Africa has a long history of exploiting *D. hottentotus*, the fishery in Angola is only in its infancy. Fishery pressure has been linked to genetic erosion in other taxa (e.g. McKeown *et al*., 2017) and may have contributed to the lower levels of mtDNA variability among South African *D. hottentotus*. Fine scale genetic studies in both regions are thus needed to inform both fishery sustainability and preservation of adaptive potential (Iles and Sinclair, 1982; Ryman *et al*., 1995; Ruzzante *et al*., 2006; Therkildsen *et al*., 2013).

4.4.3 Within-region panmixia

Microsatellite variation revealed no evidence of differentiation among the two South African samples. Power analysis indicates that the microsatellite data set has a high resolving power for detecting low levels of population structuring being >95% for F_{ST} values ≥ 0.0125. Whilst adult *D. cervinus* are known to exhibit residence behaviour, and so may not disperse far once settled, the species has a pelagic larval phase that could facilitate long distance dispersal and gene flow assuming favourable habitat conditions (Macpherson and Raventos, 2006). Larval dispersal has been proposed as a mechanism underpinning dispersal and a consequent lack of structuring within Angolan and South African waters for a number of other fish with long PLD's such as *Lichia amia, Argyrosomus japonicus, Argyrosomus coronus, Atractoscion aequidens* and *Diplodus capensis* (Henriques, 2012). This study only includes two sampling sites from South Africa in the microsatellite analysis so it cannot be ruled out that given more sampling sites and greater geographical distances between sites in South Africa some finer level of genetic structuring would not be observed. The difficulties of deriving quantitative estimates of gene flow and dispersal from subtle genetic structure among large populations (Whitlock and McCauley, 1999; Palsboll *et al*., 2007; Hellberg, 2009), and discrepancy between levels of gene flow needed to limit genetic differentiation and

dispersal to replenish stocks (Hauser and Carvalho, 2008) are fundamental issues. Therefore, while the low level of genetic structure within regions for various species is compatible with high gene flow, it cannot be ruled out that there is significant isolation of stocks on timescales of interest to management. Resolution of such spatial stock structure may be beyond the level of neutral genetic markers and would benefit from complementary analysis of markers under selection (Canino *et al*., 2005). Additionally, spatial structuring may be influenced by local species / environment interactions (Banks *et al*., 2007; McKeown *et al*., 2017). Therefore, the lack of structure among *D. hottentotus* must not be used to infer a similar lack of structure among populations within Angolan waters.

4.4.4 Historical demography

Mismatch distribution analyses revealed signatures of historical demographic fluctuations among the Angolan and South African samples. Other demographic tests (Fu's *Fs* and Tajima's *D*) were negative for both values in both regions, however Fu's *Fs* was only significant in Angola and Tajima's *D* was only significant in South Africa (Table 4.4), although the power of these tests is likely limited by the low levels of polymorphism (Kašparová *et al.,* 2015). Such low levels of variability have also likely contributed to the wide confidence intervals associated with the observed τ values (Schneider and Excoffier 1999; Kašparová *et al.*, 2015). Despite these wide confidence intervals mean τ estimates obtained from the various methods revealed a congruent trend. All estimates of time since expansion agreed that there has been a more recent population expansion in Angola and an older expansion in South Africa. Given the wide range of estimates of the τ value, dates for time since expansion for South African *D. hottentotus* range from 250 Kya (Arlequin) to 67 Kya (SITES), likewise Angolan *D. hottentotus* times since expansion range from 102 Kya (Arlequin) to the present day (DnaSP). Such estimates of time since expansion broadly coincide with the previous (Ipswichian) interglacial for both South African and Angolan populations, although the data cannot rule out a more recent expansion in Angola during the present interglacial. Other coastal fishes in the region display signals of population expansions dating from the last glacial maximum: ~27 Kya in Angolan and ~25 Kya in South African populations of *Atractoscion aequidens* (Henriques *et al.,* 2014); ~31 Kya in Angolan *Argyrosomus coronus* (Henriques, 2013); and 18 Kya in Angolan and ~13 Kya in South African populations of *Lichia amia* (Henriques *et al.,* 2012). The closely related *D.*

capensis exhibited an older time since expansion in South Africa (~40 Kya) and a more recent expansion in Angola (~8 Kya), agreeing with the present study for *D. hottentotus* in pattern if not the detail of timescales. However, the use of the more conserved COI sequence, with its limited variation in these *D. hottentotus* populations, may have limited the resolution able to be achieved (Grant, 2015), and so it is possible that *D. hottentotus* also may have undergone population expansions on a similar timescale to other coastal fish in the region, i.e. during and since the last glacial period.

5.5 Conclusions

This study identifies genetic divergence between populations of *Diplodus hottentotus* in southern Africa at a level similar to that seen between these and Mediterranean populations of *D. cervinus*, which does not coincide with the proposed species designation of *D. cervinus* in the North East Atlantic and Mediterranean and *D. hottentotus* in Angola and South Africa. The Angolan – South African division is supported by morphological and life history divergence investigated by Winkler (2013). On balance, the data available suggest that the *D. hottentotus* population in Angola should be recognised as divergent from the South African population by appropriate taxonomic recognition. As such the present study highlights that DNA barcoding has great value as an exploratory technique in taxonomy and for revealing cryptic diversity. However, it also shows that this potential can only be maximised if traditional COI based approaches are complemented with data from other (independent) genetic loci and ontogenetic data. In light of the dynamics of speciation in the region, failure to do so or reliance on one method may compromise species delimitation and an underestimation of coastal southern African ichthyodiversity, thereby curtailing efforts to conserve evolutionarily distinct taxa in this intriguing and complex marine system. Regarding recurrent management strategies, the lack of intra-regional genetic structure must not be assumed to reflect a lack of biological stock structure, and until more comprehensive studies are carried out (more markers and cohort-specific sampling) a spatial bet hedging approach is recommended.

Chapter 5: Synthesis

5.1 Overview

The present study has identified a range of phylogeographic scenarios displayed in coastal fishes along the west coast of Africa, a fundamental finding being that the Benguela Current System represents a major historical and recurrent barrier to gene flow that has profoundly shaped biodiversity in the region. mtDNA genotyping of samples across the Benguela Current system revealed reciprocal monophyly of clades within some populations on either side, and shallower levels of genetic divergence between other populations compatible with incomplete lineage sorting, as well as signatures of intermittent secondary contact in several species. Similar phylogeographic scenarios have been suggested from previous work on fish living within the Benguela Current LME (Henriques 2012; Henriques *et al*., 2012, 2014, 2016), and also in the few studies which have studied coastal fish phylogeography along the west coast of Africa (Durand *et al*., 2005, 2013). It is therefore pertinent to not only summarise the present study but also to collate and identify common overarching models of population connectivity (both present day and historical), as well as historical population crashes and expansions in coastal fish along the Atlantic coast of Africa. The pertinent details of this study and other relevant phylogeographic studies along the west coast of Africa are summarised in Table 5.1. Upon studying the collated evolutionary histories five key events can be established:

1. *Population Divergence during the Pliocene / Pleistocene Transition*. The oldest genetic divergences detected in the present study are associated with the Atlantic-Mediterranean transition. The date of divergence for Atlantic and Mediterranean lineages of *Spondyliosoma* (2.25- 2.58 Ma) coincides with the proposed environmental destabilisations and habitat fragmentation occurring in the western Mediterranean after the onset of glacial / interglacial periods at the Pliocene / Pleistocene transition 2.6 Ma (Patarnello *et al*., 2007). In contrast *Lithognathus mormyrus* exhibits a divergence between Atlantic and Mediterranean lineages dating approximately from 3.24-3.69 Ma, a million years older than the implied habitat fragmentation between the Atlantic Ocean and Mediterranean Sea at the Pliocene / Pleistocene transition. The Sala-Bozano *et al*. (2009) study suggest the initial divergence between Mediterranean and Atlantic *L.*

mormyrus resulted from the recolonisation of the Mediterranean following the Messinian Salinity Crisis, when conditions in the region stabilised around 3 Ma. Mediterranean *L. mormyrus* individuals then subsequently expanded their range into the Atlantic resulting in the observed secondary contact in the Alboran Sea found by Sala-Bozano *et al*. (2009). However given higher 'resolution' CR sequence data in the present study we identify that the present day North East Atlantic Clade III diverged from their Clade I ancestors approximately 370 Kya. As such the present day observed Atlantic and Mediterranean clades would have resulted from a North- South divide in the Atlantic 3 Ma and the subsequent more recent recolonisation of the North East Atlantic by austral *L. mormyrus* individuals. This suggests an even older role for the biogeographical barriers presented by the Benguela Current and Tropical barriers, both of which were present 3 Ma (Siesser, 1980; Diester-Haass *et al*., 1988; Diester-Haass *et al.*, 1990; Krammer *et al*., 2006).

Table 5.1 Collation of results from population genetic studies focussing on the species distributed around the Benguela Current region. Parameters: Larvae = larvae pelagic or demersal; Sex Change? - is the species hermaphroditic; Tropical Barrier = evidence of breakdown of gene flow between Angola and North East Atlantic samples; BC = evidence of breakdown of gene flow across the Benguela Current; tmrca = time since most recent common ancestor; Speciation? = any evidence the species has undergone speciation across the BC; Morphology? = any evidence of morphological divergence across the BC; Texp = Time since expansion (ZA = South Africa; ANG = Angola). F_{ST} = reported F_{ST} values (and their equivalent) across BC (Msat= Microsatellite); References: *A. aequidens*: Henriques *et al*. (2016); *Argyrosomus*: Griffiths and Heemstra, (1995); Henriques, (2012); *L. amia*: Henriques *et al*. (2012); *O. vulgaris*: de Beer, (2014); *D. capensis*: Richardson (2010); Henriques (2012); *T. megalopterus*: Soekoe, (2016) and *S. vermiculata*: Healey *et al*., (2017)..

Species	Larvae	Sex Change?	Tropical Barrier?	BC Barrier?	Tmrca	Speciation?	Morphology	Texp ZA Kya	Texp ANG Kya	F_{ST} BC
Atractoscion aequidens	Pelagic	Gonochoristic	-	YES	2.00 Ma	YES	NO	24.53 Kya	27.41 Kya	CR Φ_{ST} = 0.902 Msat F_{ST}- 0.055
Argyrosmus spp.	Pelagic	Gonochoristic	YES	YES	2.00 Ma	YES	YES	30.7 Kya	-	-
Spondyliosoma spp.	Demersal	Protogynous	YES	YES	1.92 Ma	YES	YES	42 Kya	26 Kya	COI Φ_{ST}- 0.966; CR Φ_{ST}- 0.956; Msat F_{ST}- 0.094- 0.103
Lithognathus mormyrus	Pelagic	Protandrous	YES	YES	786 Kya	NO	NO	174 Kya	259 Kya	COI Φ_{ST}- 0.220; CR Φ_{ST}- 0.092; Msat F_{ST}-0.025-0.039
Sarpa salpa	Pelagic	Protandrous	YES	YES	708 Kya	NO	NO	74 Kya	161 Kya	COI Φ_{ST}- 0.304; CR Φ_{ST}- 0.444; Msat F_{ST}- 0.028- 0.031
Lichia amia	Pelagic	Gonochoristic	-	YES	202 Kya	NO	-	12.69 Kya	17.76 Kya	CR Φ_{ST}= 0.78
Octopus vulgaris	Pelagic	Gonochoristic	-	YES	231Kya-1 Ma	NO	NO	129 Kya	100Kya	Cytb F_{ST}- 0.682-0.729
Diplodus capensis	Pelagic	Protandrous	-	YES	367 Kya	YES	YES	40.38 Kya	8.14 Kya	COI F_{ST}- 0.832-0.950 Msats F_{ST}- 0.039- 0.048
Diplodus cervinus/ hottentotus	Pelagic	Protogynous	-	YES	-	YES	Some	200 Kya	82 Kya	COI Φ_{ST}- 0.496 Msat F_{ST}- 0.236-0.284
Triakis megalopterus	Ovoviviparous	Gonochoristic	-	YES	-	NO	Some	-	-	CR F_{ST}- 0.150- 0.717 Msat F_{ST}- 0.00- 0.112
Sepia vermiculata	Demersal	Gonochoristic	-	YES	-	NO	-	-	-	-

2 *Benguela and Tropical barriers to dispersal drive cryptic and non-cryptic speciation ~ 2 Ma.* The next oldest divergences are observed between allopatric populations either side of the Benguela Current region, with dates of divergence falling around 2 Ma: *Spondyliosoma* (1.92-2.13 Ma, this study); *Atractoscion aequidens* (2 Ma, Henriques *et al*. 2016); *Argyrosomus japonicus* / *Argyrosomus coronus* (2 Ma, Henriques, 2012). Around 2 Ma the Benguela Current assumed its present day conditions (i.e. establishment of the perennial Lüderitz upwelling cell and subsequent cooling - Marlow *et al*., 2000, Krammer *et al*., 2006) which led to complete and sustained disruption of gene flow between Angolan and South African populations of *Spondyliosoma, At. aequidens* and *Ar. japonicus*/ *Ar. coronus* resulting in genetic divergence associated with speciation in these taxa. Seemingly concurrent to the allopatric divergences associated with the intensification of the Benguela Current, populations of warm temperate coastal fishes in Angola and the North East Atlantic were isolated by the Tropical Barrier. Again this appears to have led to allopatric speciation in *Ar. coronus* (Angola) and *Ar. regius* (North East Atlantic and Mediterranean; Henriques, 2012), and probable cryptic speciation in *Spondyliosoma* (1.67- 2.28 Ma, this study). Observed genetic divergences in the *Sepia officinalis* species complex (Healey *et al*., 2017) are also likely to be associated with these barriers and timescales. The results of this study and that of Henriques (2012) identify that not only are the Tropical equatorial currents a barrier to dispersal in warm temperate coastal fishes, but that these currents are also as impermeable as the Benguela Current for some coastal fishes.

3 *Early Pleistocene permeability and Mid-Pleistocene Transition divergence.* Both *L. mormyrus* and *Sarpa salpa* exhibit dates of initial genetic divergences between populations across the Tropical and Benguela Current Barriers during the final stages of the Mid-Pleistocene Transition (MPT). For *S. salpa* the tropical barrier between the North East Atlantic and Angola was permeable up to the MPT. However following the onset of longer and more pronounced glacial cycles from 900 Kya onwards the tropical barrier became impermeable for *S. salpa*, with a date of divergence between Clades III (North East Atlantic) and II (Angola) of 853 Kya.

The Benguela Current was permeable to both *L. mormyrus* and *S. salpa* during the MPT. During the MPT the Benguela Current was (relatively) warmer and of reduced intensity and generally less stable (Marlow *et al*., 2000; Clark *et al*., 2006). These conditions may have allowed for occasional dispersal from south to north (for both

adults and larvae) as well as permitting southward dispersal of adults, but the northward flow of the Benguela Current making it an unlikely route (south) for larval dispersal. Of particular note is that *S. salpa* appears to be a rare example of an Atlantic coastal fish colonising the Indian Ocean, whereas most colonisations are found to be from the Indian Ocean into the Atlantic Ocean (Rocha *et al*., 2005; Floeter *et al*., 2008). After the end of the MPT the Benguela Current Barrier drove divergence in both species, which exhibit genetic divergences between Angola and South Africa dating from 786 Kya (*L. mormyrus*) and 708 Kya (*S. salpa*).

4 *Population expansions and secondary contact*. Phylogenetic analyses suggest that around 370 Kya *L. mormyrus* Clade III individuals colonised the North East Atlantic, diverging from its Clade I ancestors in South Africa. IMa2 analysis suggests asymmetrical secondary contact between clades I and II in Angola for both *L. mormyrus* (455 Kya) and *S. salpa* (262 Kya) across the Benguela Current. The processes driving these historical 'colonisations' is unclear. Palaeoceanographic studies of the Benguela Current identify stability in the system after MPT with the only observed changes being larger and increased upwelling activity in the Western Cape region in South Africa, as well as off Lüderitz, during glacial periods (Petrick *et al*., 2015). So rather than instability in the Benguela Current allowing occasional transport through the barrier (as suggested for above divergences dating to the MPT), another intermittent indirect passage for migrants must be present to allow for these colonisations. As suggested in Chapter 3 leakage of Agulhas warm water into the southern Atlantic during interglacial periods (Peeters *et al*., 2004; Lutjeharms, 2006; Beal *et al*., 2011) may pose the most parsimonious route. As such Agulhas leakage may play a prominent role in maintaining sporadic connectivity and promoting colonisation of the Atlantic Ocean by South African warm temperate species. Colonisation of the Atlantic (by Indian Ocean fishes) via Agulhas leakage has been previously suggested for Gobies (Rocha *et al*., 2005), Pygmy Angelfishes (Bowen *et al*., 2006) and the scalloped hammerhead shark (Duncan *et al*., 2006). Briggs and Bowen (2013) identify these colonisation events as being 'pulses' which correspond to interglacial periods. Such a scenario of colonisation via Agulhas leakage during interglacial's is also plausible for the observed more recent divergences between South African and Angolan *Diplodus capensis* (367 Kya) and *Lichia amia* (202 Kya, Henriques *et al*. 2012).

From 260 Kya onwards several study species (*L. mormyrus*, *S. salpa* and *D. cervinus*) as well as *Octopus vulgaris* exhibit population expansions (Table 5.1). Whilst these

population expansions all have considerable variability around their associated τ- value estimates they all roughly coincide with the Ipswichian interglacial (approximately 130-115 Kya), although the dates of expansion for the Angolan *L. mormyrus* population could date to older interglacial's during the Wolstonian. During interglacial periods favourable conditions for warm temperate species would be more prevalent, with increased available habitat due to higher sea levels and increased productivity (Peeters *et al*., 2004). For both *L. mormyrus* and *S. salpa* the Angolan Clade II exhibits older population expansions than their Clade I counterparts. In South Africa during cold glacial periods available coastal habitat is greatly reduced, with the lower sea levels exposing large tracts of the Agulhas Bank (Ramsay and Cooper, 2002). During glacial periods the Agulhas Current was significantly reduced in both size and intensity (Peeters *et al*., 2004) likely leading to range retractions north-eastward along the coast of South Africa in tropical and warm temperate coastal fishes such as *L. mormyrus* and *S. salpa,* leading to population contractions. Such range expansions are well documented in South African marine taxa from both genetic studies (Teske *et al*., 2011; Teske *et al*., 2013) and the fossil record of marine molluscs (Kensley, 1985). During interglacial periods the Agulhas Current increased in strength and scope along with higher sea levels covering the Agulhas Bank in South Africa (Peeters *et al*., 2004), leading to an increase in available coastal habitat for fishes such as *S. salpa* and *L. mormyrus* promoting population expansions. As a cautionary note such dates of population expansions (i.e. pre LGM) are controversial given the limitations of dating historical population expansions (Grant, 2015); it remains however likely that population expansions occurred during previous interglacial periods even if it is contentious that such patterns would be revealed in contemporary genetic demographic signals.

5 *Population expansions since the last glacial maximum*. Many species exhibit population expansions since the last glacial period, such as *Spondyliosoma, At. aequidens, D. capensis* and *L. amia* (Table 5.1). Such proposed expansions are logical, as since the last glacial maximum sea levels have risen leading to increased available habitat and overall increases in marine productivity in the southern African region (Pirazzoli, 1997). What perhaps is most striking is that both the timing of these population expansions (during interglacials) and geographical position (Agulhas Bank, South Africa) may have increased the potential for northward colonisation via Agulhas leakage. Population increases during the present (and probably historical) interglacial

would place an abundance of larvae in the exact geographical location (Agulhas Bank) where Agulhas rings peel off the Agulhas Current retroflection and into the south Atlantic (Peeters *et al*., 2003), both essentially acting together to increase the probability of colonisation by this route. As such colonisation of Angola (by both *S. salpa* and *L. mormyrus*) and the North East Atlantic by *L. mormyrus* may have been indirectly driven by historical population expansions in South Africa.

5.2 Angola as a region of endemism: implications for conservation and fisheries.

The warm temperate and subtropical coastal regions of northern Namibia and southern Angola harbour a high level of both intra- and inter-specific endemism in warm temperate coastal fish fauna (Potts *et al*., 2015) (Table 5.1). This endemism is promoted by being bounded by two biogeographical barriers, the warm equatorial currents to the north and the cold Benguela Current to the south. The west coast of Africa is characterised as being depauperate compared to other marine biogeographic regions for marine biodiversity (particularly tropical fishes; Floeter *et al*., 2008). However the results of the present study and those before it strongly indicate that the past and present day oceanographic features have and are promoting ongoing speciation in warm temperate taxa and thus contributing to marine biodiversity. Whilst divergent Angolan clades in many of these species may not conform to species level they do satisfy the criteria for being considered as Evolutionary Significant Units (ESUs; Moritz, 1994).

The finding that Angola harbours unique genetic diversity has significant implications for conserving and managing these evolutionarily distinct fishes. In the present study *Spondyliosoma*, *L. mormyrus*, *S. salpa* and *D. cervinus* cannot be considered as trans-boundary stocks, i.e. a single stock shared between Angola and South Africa, and should not be managed as one. Whilst the present study identifies limited evidence of population structuring within South African and Angolan waters for the study species this should not be confused or used to infer that there is not isolation of stocks on timescales of interest to fisheries management. Such fine scale structuring may be beyond neutral genetic markers utilised here, and markers under selection may be required to elucidate such structuring (Canino *et al*., 2005).Secondly the present study focused on the Cape region of South Africa, lacking samples from the east coast (e.g. KwaZulu-Natal) where several biogeographic boundaries and divergent lineages have

been previously identified (Teske *et al*., 2007; Sala-Bozano *et al*., 2009; Teske *et al*., 2011).

In Angola after decades of civil war the fisheries are not as developed as they were historically or compared to other fisheries in the region (such as South Africa; Potts *et al*., 2009).However the subsistence inshore fishery in Angola has become more exploited since the end of the civil war (Potts *et al*., 2009), with some fish (e.g. Cunene Horse Mackerel, *Trachurus trecae*) now considered to be overfished (FAO, 2014). So whilst the region is not unexploited, Angolan fisheries are still in development and as such can be enabled for sustainable exploitation of fishery stocks. Marine Protected Areas (MPAs) may go some way to help conserve the endemic marine fauna in the region as well as allowing for their sustainable exploitation (Perez-Ruzafa *et al*., 2006; Edgar *et al*., 2007; Arrieta *et al*., 2010; Briggs, 2011). MPAs are protected regions which are unexploited leading (in theory) to a healthy population of fish which will then disperse out of the protected region thereby replenishing exploited populations. However as a cautionary note MPAs only function well when they are well designed and protected. Edgar *et al*. (2014) in their global study of MPAs identified five key factors that work well, namely full protection (i.e. not exploited commercially or recreationally), enforcement of that protection, age of MPA (>10 years to attain benefits of MPA), size (best > 100km^2) and in isolated locations (e.g. isolated reefs). Edgar *et al*. (2014) identified that at least three of these factors are required for an MPA to have any conservation value, with the best marine reserves attaining all five factors. Some mitigation by loss of such factors may be possible by factoring in connectivity between MPAs (Halpern, 2014). For any MPAs proposed for the southern African region it would be highly desirable that they have the above attributes. However for MPAs to be truly effective they need to cater not only to conservation needs but also to needs of local communities and fishermen, indeed involvement of local fishermen is imperative for successful MPAs (Di Franco *et al*., 2016). In Angola local fisher cooperatives as well commercial fisheries would need to be involved in all aspects of any implementation of MPAs in the region.

5.3 Dispersal potential

Life history traits are often suggested to correlate to population genetic structuring and speciation events. Of particular interest in marine fishes are egg type (demersal / pelagic), whether the larvae spend time in the water column, and migratory ability; all

traits which could profoundly affect realised dispersal and levels of population structuring (Galarza *et al*., 2009). Therefore it is feasible that some traits may predispose species to higher levels of population genetic divergence associated with the two major biogeographic boundaries on the west coast of Africa.

Of the species studied here only *Spondyliosoma* exhibits a demersal egg type and larval stages that do not spend time in the water column. *Spondyliosoma* also exhibits the most sensitivity to the key biogeographic barriers identified in the present study with the Mediterranean-Atlantic transition, Tropical Barrier and Benguela Current Barrier all being associated with proposed cryptic speciation events. However the congeneric *Ar. japonicus*, *Ar. coronus* and *Ar. regius* all have pelagic spawning and a pelagic larval phase (i.e. spends time in the water column), as does *At. aequidens*, and yet the Benguela Current and Tropical Barriers appears likewise to have promoted allopatric speciation in these taxa (Henriques, 2012; Henriques *et al*., 2016), an unexpected finding if species-level genetic divergence across the Benguela Current were driven by egg and larval type. Whilst the remaining species listed in Table 5.1 do not display species-level divergence they all display significant genetic differentiation across the Benguela Current and display pelagic spawning and have larvae which have a pelagic larval phase (with the exception of the smooth hound shark, *Triakis megalopterus*, which is ovoviviparous). As such there seems to be no corroboration that genetic divergence across the Benguela Current and egg / larval type.

Whilst dispersal potential in the larval stages does not correlate to the observed population genetic structuring in fish species across the study region, what of adult dispersal potential? Both *Spondyliosoma* and *Argyrosomus* spp. undergo spawning migrations as adults, indicating a high level of dispersal potential (Griffiths and Hecht, 1995; Pawson, 1995). However both *Spondyliosoma* and *Argyrosomus* spp. exhibit species-level divergence across the Benguela Current and Tropical barriers. Likewise *L. mormyrus* and *S. salpa* exhibit similar levels and timings of genetic divergence across these biogeographic barriers regardless that *S. salpa* undergoes migrations for spawning (i.e. high adult dispersal potential) and *L. mormyrus* adults are generally sedentary (i.e. low adult dispersal potential). Tagging studies of *Argyrosomus* have identified Angolan individuals in North West South Africa indicating that adults can traverse the Benguela Current, yet genetic divergence between the two regions is maintained indicating Angolan migrants do not breed with South African individuals (W. Potts, pers. comm.).

Indeed *Argyrosomus* exhibits philopatric migrations within South Africa, indicating that for some species even if adults can readily disperse through major biogeographic barriers, gene flow may still not be possible as individuals need to return to their natal breeding grounds to spawn. Further studies are needed of the present study species life history (especially PLD) and physiological tolerances to fully evaluate the relationship of any correlation between life history traits and observed genetic population structuring. But collectively the present studies so far (as outlined in Table 5,1) indicate genetic divergence is driven by the seascape features of the Benguela Current region and less so by intrinsic species properties.

5.4 Effects of life history characteristics: Sex change

As outlined in the Introduction, hermaphroditism may have an impact on observed population structuring by affecting N_e. Since all species in the present study are hermaphrodites it is prudent to investigate and compare the levels of genetic population structuring observed in other species that are not hermaphrodites (gonochoristic-see Table 5.1) and sequential hermaphrodites, and then secondly between protandrous and protogynous hermaphrodite species. Perhaps of most significance in this study is the effect of hermaphroditism on the N_e of the mitochondrial markers. MtDNA is maternally inherited, and generally accepted to have a N_e ¼ that of the nuclear genome. However in sequential hermaphrodites every individual can potentially contribute mtDNA to the next generation meaning theoretically it can be as high as ½ of the N_e of the nuclear genome The N_e is further complicated by unequal sex ratios in hermaphrodites. In a strict logical sense for protandric species (*S. salpa* and *L. mormyrus*) we would expect an overall reduced mtDNA N_e leading to an increased level of genetic structuring. Whereas for our two protogynous species *Spondyliosoma* and *D. cervinus* we would expect a larger mtDNA N_e, leading to less genetic drift and a lower level of genetic population structuring (Coscia *et al*., 2016). As such we can formulate some broad predictions regarding sex change and the amount of genetic structuring observed across the major barriers to dispersal for coastal fishes. Firstly, that hermaphroditic species will exhibit a higher level of genetic population structuring than their gonochoristic counterparts. Secondly that protandrous species will exhibit a higher level of genetic structuring in comparison to protogynous species.

Looking exclusively at the Benguela Current Barrier there does not appear to be any observable correlation between increased genetic population structuring (i.e. larger

divergences) and hermaphroditism. *Spondyliosoma* which is a protogynous hermaphrodite exhibits species level divergences whilst the gonochoristic *Argyrosomus* spp and *At. aequidens* also exhibit speciation events associated with the Benguela Barrier. Such results suggest that any role of hermaphroditism increasing genetic population structuring is secondary to other biological traits, chance and palaeoceanography. However interpreting such a result should be treated with caution, these are very broad comparisons between fishes of different genera and families with a low level of phylogenetic relatedness. Ideally such comparisons would need to be made between congeneric fish species where one species is gonochoristic and another hermaphroditic to reduce the impact of phylogenetic distance.

Similar to the above gonochoristic-hermaphroditic comparison the protandric / protogynous comparisons yielded no clear correlation to an increased level of genetic structuring associated with protandrous species. *Spondyliosoma* as a protogynous fish would be expected to show the lowest levels of divergence across the Benguela Current Barrier, yet exhibits a cryptic speciation event. Indeed *Spondyliosoma* exhibits cryptic speciation events across all the major biogeographic barriers in this study; whilst the two protandrous species, *L. mormyrus* and *S. salpa*, exhibit shallower levels of genetic divergence across the Benguela Current. Once more these results appear to run contrary to the predictions of a higher level of genetic population structuring in hermaphrodites. Again this could be due to confounding factors such as dispersal ability, territoriality and other life history traits. These rather broad conclusions are questionable due to the relatively low level of phylogenetic relatedness between the study species, although they all are at least within the same family (Sparidae).

This potential issue of phylogenetic relatedness however can be tested directly with *Diplodus cervinus* and *Diplodus capensis*. These species are closely related (Summerer and Sturmbauer, 2001; Santini *et al*., 2014), and both hermaphrodites with *D. cervinus* being protogynous (Pajuelo and Lorenzo 2001, Pajuelo *et al.,* 2003a, 2003b) and *D. capensis* being protandrous (Mann and Buxton 1998; Richardson, 2010) allowing for a direct comparison between two sister species. In this comparison the two species observed genetic divergences across the Benguela current conforms to the above hypothesis. The protandrous *D. capensis* exhibits a pronounced genetic divergence in the COI gene with three fixed differences and no shared haplotypes between Angolan and South African fish (Table 5.1; Henriques, 2012). The protogynous *D. cervinus* however exhibits no observed fixed differences across the Benguela Current and also

has shared haplotypes across both regions. This is important as it shows that perhaps protandrous species do exhibit an increased level of population genetic structuring across biogeographic barriers in mitochondrial genetic markers. However this result should again be treated with some level of caution since a significant level of population differentiation between Angolan and South African *D. capensis* is observed by Henriques (2012) utilising nuclear microsatellite loci (i.e. unaffected by reduced mtDNA N_e). Also the difference in levels of divergence across the Benguela Current in *D. cervinus* and *D. capensis* could also be due to other life history traits as outlined in Chapter 4. This is also a single result and further species pairs would need to be examined to establish if this is true more broadly in hermaphroditic fish.

Overall, the evidence for a role of hermaphroditism in increasing observed genetic population structuring is limited when comparing distantly related species. Whilst at the family level there is no clear correlation to an increased level of population structuring in protandrous fish species, although further Sparids should be included. Whilst at the level of sister species this study does find some support for the theoretical expectations that a protandric species would exhibit a higher level of genetic structuring (as observed by mitochondrial markers) than its congeneric protandric species.

5.5 Comparison to other marine biogeographic barriers.

In the marine realm levels of genetic population structuring are presumed to be small (Waples, 1998; Hauser and Carvalho, 2008). This is true for many widespread pelagic species which are not restricted to the shallow continental shelf (as species in the present study are); for example Bluefin Tuna (*Thunnus thynnus*; mtDNA F_{ST}= 0.023; nDNA F_{ST}= 0.002 – Carlsson *et al*., 2004) and Bigeye Tuna (*Thunnus obesus*; nDNA F_{ST}= 0.000 to 0.003 - Gonzalez *et al*., 2008) have relatively shallow population structures on an oceanic scale. Such structures are maintained by (at least historically) large population sizes and the ability for pelagic individuals to disperse and breed freely throughout the open ocean. However this pattern is not common for coastal fishes such as the Sparids in the present study. In coastal and reef-associated fishes population structure is governed by oceanographic features such as land bridges, upwelling's, frontal systems and large stretches of Open Ocean. Barriers to dispersal (and gene flow) in the marine realm are often referred to as being 'hard' or 'soft' (Cowman and Bellwood, 2013). A land barrier such as the Isthmus of Panama is considered to be a hard barrier, whilst features such as frontal systems, currents and the open ocean are

considered as soft barriers since it is physically possible for taxa to disperse through such barriers. The Benguela Current and the tropical equatorial currents forming the Tropical Barrier are thus considered to be 'soft barriers'. It is clear from the present study and those before it that the Benguela Current presents a formidable barrier to dispersal in both the present day and historically. It is thus of interest for us to compare the Benguela Current barrier to other biogeographic barriers in the marine realm, both hard and soft.

A hard barrier: *the terminal Tethyan event*

The closure of the Tethys Sea occurred approximately 13-18 Ma. Unsurprisingly given the age of such a barrier its effects are seen at the family level rather than the species level. In a meta-analysis by Cowman and Bellwood (2013) studying the three families Pomacentridae, Labridae and Chaetodontidae, vicariant events were identified in all three families dating to many millions of years before the final closure and the barrier presented by the Arabian Land Bridge. The closure of the Tethys Sea took many millions of years over the middle of the Cenozoic. This resulted in the fragmentation of much of the palaeomarine Tethys habitat, leading to population fragmentation and divergences between Atlantic and Indian Ocean marine taxa predating the final closure of the Tethys. Given the age of the closure of the Tethys it is perhaps not surprising that the levels of divergences are of much greater magnitude than seen across the Benguela Current. Yet even in this formidable land barrier there is an indication of habitat fragmentation and palaeoceanographic processes contributing to the observed divergences rather than continental drift and hard land barriers acting alone.

A hard barrier: uplift of the Isthmus of Panama.

The uplift of the Isthmus of Panama and subsequent separation of the Pacific and Atlantic fauna was fully complete by 2.8Ma, although the whole process took over 15 Myrs (Coates and Stallard, 2013), or even longer at 24 Myrs (Bacon *et al*., 2015; Marko *et al*., 2015). This date of closure of the land bridge presents a definitive date to population separation of species pairs (geminate) from their common ancestor. Lessios's (2008) review of divergence of marine organisms after the rise of the Isthmus of Panama finds the approximate 3 Myrs of divergence is enough for both genetic and behavioural divergence in echinoids, crustaceans, fishes and molluscs but seemingly not enough for complete reproductive isolation, finding only complete reproductive isolation in *Diadema*. The Lessios (2008) review has implications for the present study: given that the Benguela Current system (in its present form) has been in place for less

time than the Isthmus of Panama it is unlikely that diverged lineages which subsequently come into secondary contact would show complete reproductive isolation (i.e. have undergone complete speciation forming good biological species) since without ecological factors this process can take in excess of 3 Myrs to complete (Coyne and Orr, 1997; Lessios, 2008). Similar to the terminal Tethyan event, vicariance in marine taxa across the Isthmus of Panama is dated to up to 12 Ma, with an increase in occurrences of vicariance reaching a crescendo just before the formation of the hard land barrier (Lessios, 2008; Cowman and Bellwood, 2013). These vicariant events were most likely driven by habitat fragmentation during the geological processes driving the uplift of the Isthmus of Panama. Again this shows the importance of habitat change and loss in driving vicariant events even before hard physical land barriers form. Perhaps of relevance to the Benguela Current Barrier is that although it is not a 'hard' barrier, for some taxa it has been impermeable since the onset of present day conditions 2 Ma. The Benguela Current has also been in place for up to 10 Myrs resulting in habitat loss and fragmentation for both warm temperate and tropical coastal fishes. It may also prove to be fruitful to examine the impact of the Benguela Current at the level of genera or even family.

A soft barrier: mouth of the Amazon River.

The Amazon River originated as a pan-continental river around 11 million years ago and took its present shape in the late Pliocene around 2.4 Ma (Figueiredo *et al.*, 2009). The freshwater outflow offshore from the Amazon presents a dispersal barrier to coastal marine reef fishes. However estimated times of divergence of taxa on either side appear to be much younger than the age of onset of the Amazon river outflow, with the oldest known divergence found in *Ophioblennius macclurie trinidadensis* at 6 Ma based on mtDNA Cytb (Muss *et al*., 2001), to 1 Ma in *Acanthurus bahianus* (Rocha *et al*., 2002), to very recent in *Acanthurus coeruleus* (Rocha *et al.*, 2002). These findings are attributed to the Amazon barrier being 'soft', that is to say it is sometimes passable for marine fishes during glacial periods when outflow is lowered and the species life history attributes, particularly their tolerance to freshwater and habitat preference, allow for some dispersal. The *Acanthurus* species *A. bahianus* occurs on rocky reefs which are absent in the Amazon mouth region whist *A. coeruleus* can tolerate softer bottom substrate (hence the recent divergence), whilst *A. chirurgus* can happily tolerate soft substrate and shows no divergence associated with the Amazon barrier. These finding appear to mimic the findings in phylogeography of West African coastal fishes whereby

we find ancient divergences associated with the cold Benguela Current in *Spondyliosoma*, *Argyrosomus* and *At. aequidens* (all diverging approximately 2 Ma) to those associated with more recent interglacial periods in *L. amia* (202 Kya, Henriques *et al*., 2012). This indicates the Benguela Current system can act as a porous barrier, just like the Amazon, allowing some fishes to disperse more freely than others. Recently the discovery of an extensive reef system at the mouth of the Amazon River may allow for alternative modes of dispersal across the Amazon barrier (Moura *et al*., 2016). The work on *Acanthurus* (Rocha *et al.*, 2002) is of importance to the present study since it shows that the presence and distribution of suitable habitat may be more important than more obvious barriers such as freshwater outflow in the Amazon and the cold water upwelling in the Benguela system.

A soft barrier: the Tropical Barrier

Many marine fishes have antitropical distributions separated by the warm equatorial waters. These warm tropical waters around the equator are unsuitable for cold water adapted species, presenting a conundrum as to how such species are found to either side of such a formidable barrier to dispersal. Antitropical taxa exhibit levels of genetic divergence ranging from distinct populations to sister species (Bowen *et al*., 2016). In a review by Burridge (2002) of Pacific ocean fishes with antitropical distributions a model of divergence of such taxa by plate tectonics was ruled out, as in the Pacific most species with an antitropical distributions exhibit genetic divergences dating to the Pleistocene, with most dispersal events thought to occur during glacial periods when the tropical waters were 2-3°C cooler. However dispersal through the tropics may also be facilitated by oceanographic features. In the Atlantic, dispersal by anchovies (*Engraulis)* across the tropical barrier has been facilitated by the cold Benguela upwelling which occurs at a relatively low latitude (Grant and Bowen, 2005). In the present study all species have antitropical distributions within the eastern Atlantic. Similar to those in the Pacific Ocean some species exhibit cryptic speciation (*Spondyliosoma*) whilst other species present more recent divergences dated to the mid- to late-Pleistocene (*L. mormyrus, S. salpa* and *D. hottentotus / cervinus*).Such a pattern underlines species-specific responses to the tropical barrier similar to that identified in the review by Burridge (2002). Whilst traversing the tropical barrier during glacial periods may be possible in the eastern Atlantic it is complicated by the narrow continental shelf in the region presenting a conflict for dispersal of shallow water

coastal fishes: whilst the tropics may be cooler during glacial periods (favouring dispersal) the lower sea levels would lead to reduced habitat in the tropics and surrounding regions (limiting dispersal). Finally, the antitropical distribution is complicated in the eastern Atlantic for warm temperate taxa which also have to contend with a cold water barrier in the form of the Benguela Current; as such many species in the region exhibit isolated populations bounded by warm and cold barriers promoting taxonomic endemism between regions.

A soft barrier: the East Pacific Barrier.

The East Pacific Barrier is the widest and oldest biogeographic barrier in the marine realm. It has been in place since the late Cretaceous and separates the Indo-West Pacific from the East Pacific and Atlantic regions. The barrier is in the form of large expanses of Open Ocean unsuitable for residence or dispersal across by coastal species and is thus considered to be a 'soft' biogeographic barrier (Cowman and Bellwood, 2013). The EPB was first identified by Darwin in the Origin of Species (1872). The effectiveness of the barrier was established firstly by enumeration of shallow water species which are not shared by the two regions to either side (Ekman, 1953; Mayr, 1954 and Briggs, 1974), and more recently in molecular phylogenies where several taxa exhibit deep divergences across the EPB (Lessios *et al*., 1999, 2001; McCartney *et al*., 2000; Colborn *et al*., 2001; Collin, 2003). However like other biogeographic barriers, despite the age of the EPB there are some species which span the Pacific Ocean. Lessios and Robertson (2006) studied 20 reef fish morphospecies which are found either side of the EPB. Two fishes displayed cryptic speciation whilst amongst the other 18 species only two exhibited significant genetic differentiation across the barrier, with the remaining species exhibiting remarkable genetic similarity (Lessios and Robertson, 2008). Furthermore gene flow across the barrier was found to be unidirectional in some fishes. Breaching of the EPB is not limited exclusively to fish, with two species of sea urchin breaching the barrier (Lessios, *et al*., 1998, 2003). The scatter in the estimated times since divergence in these species across the EPB does not suggest a singular historical event which would have allowed the permeability of the EPB. Whilst biological traits such as planktonic larval duration, adult and larval physiology and dietary requirements may play a role in the observed permeability of the EPB, on such a scale and timeframe stochasticity of successful traversing of the barrier is perhaps the most prevalent factor in determining the permeability of the EPB (Lessios *et al*., 2006; Cowman and

Bellwood, 2013). As such stochasticity of successful traversing of marine biogeographic barriers cannot be ruled out (including the Benguela Current), with its importance becoming more significant for older barriers.

Concluding remarks

The Benguela Current is of comparable strength to other biogeographic barriers, regardless of whether they are hard or soft barriers. Furthermore biogeographic barriers in the marine realm do not drive divergences in just a single physical mode; rather habitat loss and change instigate initial divergences. As such the Benguela Current likely acts a barrier to dispersal in several ways; whilst the Current has been in its present form for 2 Ma, it is possible it acted as a barrier to coastal fishes prior to this in its more ancient forms (as suggested for *L. mormyrus*).

5.6 Further work and limitations

Sampling

Obtaining samples throughout a species range with a large enough number of individuals for statistical analyses is a major limitation of many phylogeographic and population genetic studies. The present study has advanced our understanding of eastern Atlantic coastal marine fish phylogeography mainly by obtaining samples from Angola, where the mid-African coastal region is still largely under described. However, the present study would have been greatly aided by additional sampling in both South Africa and Angola. Ideally a minimum of three well spaced (geographically) sampling sites in both South Africa and Angola, each with a minimum of 30 individuals per species, would have enhanced the microsatellite analyses. In South Africa further sampling from KwaZulu-Natal would have been particularly useful given the divergent lineages in *L. mormyrus* observed there by Sala-Bozano *et al.* (2009). Sampling from KwaZulu-Natal would also have allowed a more direct testing of any observable gradient from tropical to warm temperate regions in South Africa as identified in some South African marine taxa by Teske *et al*. (2011). Whilst such sampling strategies are idealised, the reality of implementing the plan is often constrained. In the present study a combination of logistics, seasonal variation in abundance of target species, funding, time constraints, and accessibility of sampling localities have limited our abilities to obtain further samples.

Outside of the southern African sampling region further sampling expeditions along the west coast of Africa (i.e. the Gulf of Guinea and Mauritania) would be required to

confirm the nature and precise locality of the Tropical Barrier. Unlike the Benguela Current Barrier where previous expeditions to Namibia have found an absence of the study species within the upwelling region itself, it remains to be confirmed whether these species are distributed along the coast of Africa from Angola to Mauritania. Species distribution maps for all of the present study species indicate a disjunct range with most species being absent from the north of Angola to the western countries of the Gulf of Guinea. The present results consequently appear to corroborate this disjunct distribution along the west coast of Africa. However such distribution maps can be limited and as such it is still possible (although unlikely) that rather than a sharp genetic split between Angolan and the North East Atlantic populations, there is a gradient between the two regions. Additionally to identify the origin of the divergent *L. mormyrus* Clade V in the KwaZulu-Natal region of South Africa further samples are required along the eastern African coast (i.e. into Mozambique) as well as from populations in Madagascar.

Genetic Markers

The present study uses neutral genetic markers, utilising both mtDNA markers and nuclear microsatellite markers to reconstruct the historical phylogeography and the present day genetic structuring. Whilst this combination allows a level of corroboration between the mitochondrial and nuclear genomes it does leave it potentially susceptible to the limitations (and advantages) of mitochondrial markers as outlined in Chapter 1 as well as being potentially complicated by the hermaphroditic nature of the study species. As such the present study would greatly benefit from utilising neutral nuclear markers such as ribosomal Internal Transcribed Spacer (ITS) regions to further corroborate the observed phylogeography of the study species. As a cautionary note, von der Heyden and Connell (2012) identified significant genetic structuring in the Sparid genus *Chrysoblephus* utilising mitochondrial markers but utilising neutral nuclear markers failed to identify any differences between species and even other genera. This could suggest that the Sparid nuclear genome is remarkably conserved and thus could underestimate population differentiation and perhaps even genera and species.

Neutral genetic markers have provided excellent insights into species genetic diversity, genetic population structuring and demographic events, in this study as well as many previous studies (Kirk and Freeland, 2011; Bowen *et al*., 2014). However adaptive genetic markers (i.e. genes which directly impact reproductive fitness) could prove to be

a fruitful area of research as a continuation of the present study. Whilst neutral genetic diversity gives a general view of the potential for a species / population to adapt to environmental change, it is the adaptive genetic diversity which will allow the malleability of a species or population to adapt to future environmental change (Hoffmann and Willi, 2008). Thus from a simple practical perspective adaptive markers would allow for better identifying and conserving those populations and species which are at most risk. Likewise identifying those genes which are promoting population divergence and speciation (Nosil and Schluter, 2011) would be very useful. Next Generation Sequencing (NGS) technologies may provide a fruitful methodology to pursue in future studies providing both neutral and adaptive markers.

NGS technologies have greatly increased the ability to quickly and cheaply analyse whole species genomes (Hemmer-Hansen *et al*., 2014). Genome wide association studies on single nucleotide polymorphisms (SNPs) may provide a means to identifying genes potentially under selection for further research. For example, Carlsson *et al*. (2013) used NGS (454) and reduced representation library construction to identify over 25,000 potential SNPs and >6000 putative microsatellite loci from ~2% of the genome of the non-model species teleost Atlantic Cod (*Gadus morhua*). Such studies develop multiple neutral and adaptive markers, as well as allowing for the potential identity and function of such genes; for example in Atlantic Cod, patterns of genetic structuring were found to be noticeably different between neutral markers and highly differentiated markers located in the candidate genes for growth and reproduction (Hemmer-Hansen *et al*., 2014). Genome data could be used in future studies (or a continuation of the present study) to assess phylogeography and patterns of gene flow across the Benguela Current. Marine phylogeographic studies incorporating NGS have been conducted on the Three-spined Stickleback (*Gasterosteus aculeatus*), identifying multiple colonisations of freshwater habitat by a panmictic marine population (Hohenlohe *et al*., 2010; Hohenlohe *et al*., 2012), in European Hake (*Merluccius merluccius*) population structure in the North East Atlantic and Mediterranean (Milano *et al*., 2014), and in European Sea Bass (*Dicentrarchus labrax*) phylogeography across the North East Atlantic and Mediterranean (Tine *et al*., 2014). The Tine *et al*. (2014) study on sea bass shows he potential for the present study, using RAD sequencing and characterising 234,000 SNPs identifying isolation of North East Atlantic and Mediterranean populations dating to approximately 270 Kya, followed by secondary contact 11 Kya

with asymmetrical gene flow from the North East Atlantic into the Mediterranean. As mentioned previously the resolution of population structuring within regions for the present study may be enhanced by utilising adaptive markers: a NGS study could identify not only neutral markers but those under selection as well. Previous studies have provided important increases in resolution for resolving fine scale population structuring in marine fishes. In Baltic Sea herring (*Clupea harengus*) subtle genetic structuring was observed between sites of less than 400 km, utilising 5989 SNPs (Corander *et al*., 2013), whilst for European Hake highly divergent outlier loci showed fine-scale structure and strong differentiation among the western, central, and eastern Mediterranean basins (Milano *et al*., 2014). Such studies show the value of applying NGS technologies to phylogeography and population management of marine taxa, and would likely be of huge benefit to future studies on the Benguela Current and tropical barriers.

Further work

Palaeoceanography. Of particular interest in the current study has been the influence of palaeoceanographic processes driving cryptic speciation, population expansions and migrations. Such associations are widely reported within the literature. It could thus be fruitful for collaboration with palaeoceanographers in future studies. This would firstly help the interpretation and likely impact of such processes on historical demography and connectivity of coastal fishes in the southeast Atlantic. Such results as identified in the present study could also be used to feedback into the reconstruction of the palaeoceanography.

Collaborators in South Africa are currently undertaking research into the potential morphological and life history differences between the South African and Angolan populations of the present study species. Such studies will contribute and inform publications resulting from this thesis. This study has however identified large divergences associated with both the Tropical and Mediterranean / Atlantic transition biogeographic barriers. As such further morphological and life history studies will be needed to substantiate the cryptic speciation that is implied in both *Spondyliosoma* and *L. mormyrus*. Of particular interest also is establishing the pelagic larval duration for *L. mormyrus, S. salpa* and *D. cervinus.* The PLD is crucial for our understanding of the feasibility of the proposed mechanisms of spread and dispersal across the west coast of Africa during previous interglacials.

5.7 Conclusion

For all species in the present study the Benguela Current acts as a major barrier to historical and present day dispersal. For *Spondyliosoma* the Benguela Current has acted as an impermeable barrier for the last 2 Myrs, whilst for *L. mormyrus*, *S. salpa* and *D. cervinus* the Benguela Current has been a semi-permeable barrier on that timescale. Most notably in *L. mormyrus* and *S. salpa* dispersal across the Benguela Current appears to have been possible during the MPT; allowing a rare Atlantic to Indian Ocean invasion by *S. salpa*. Surprisingly the warm Atlantic equatorial currents (forming the Tropical Barrier) have acted as much as an impermeable barrier to dispersal in warm temperate fishes as the Benguela Current. In *Spondyliosoma*, the Tropical Barrier has driven a further cryptic speciation event, whilst for *L. mormyrus, S. salpa* and *D. cervinus* it appears that the tropical barrier has been historically permeable during the mid to late Pleistocene. Together both the Benguela Current and Tropical Barrier have driven a hitherto unknown high level of intra- and inter-specific endemism in the warm temperate / sub-tropical northern Namibian and Angolan coasts. Whilst this study has not explicitly focussed on the Mediterranean-Atlantic transition it has nonetheless also identified this barrier as promoting cryptic speciation events in both *Spondyliosoma* and possibly *L. mormyrus*. Finally, this study identifies population expansions during interglacial periods in both South Africa and Angola. Colonisation from South Africa and secondary contact in Angola were also found in *L. mormyrus* and *S. salpa,* with the mode of transport most likely by increased Agulhas leakage into the South Atlantic during interglacial periods. As such this study highlights the complex interplay between palaeoceanography, life history and historical demography shaping historical and contemporary population connectivity in coastal fishes between the Atlantic and Indo-Pacific Oceans.

References

Alarcon, J. A. and Alvarez, M. C. (1999) Genetic identification of sparid species by isozyme markers: application to interspecific hybrids. *Aquaculture*, 173(1): 95-103.

Allsop, D. J. and West, S. A. (2004) Sex-ratio evolution in sex changing animals. *Evolution* 58(5): 1019-1027.

Almada, V. C., Oliveira, R. F., Goncalves, E. J., Almeida, A. J., Santos, R. S. and Wirtz, P. (2001) Patterns of diversity of the north-eastern Atlantic blenniid fish fauna (Pisces: Blenniidae). *Global Ecology and Biogeography* 10(4): 411-422.

Amir, S. A., Tanaka, F., Siddiqui, P. J. and Iwatsuki, Y. (2013) First records of two sparid species, Diplodus omanensis and Pagellus affinis (Perciformes: Sparidae) from western coast of Pakistan. *Cybium*, 37(3): 220-222.

Angers, B., Estoup, A. and Jarne, P. (2000) Microsatellite size homoplasy, SSCP, and population structure: a case study in the freshwater snail Bulinus truncatus. *Molecular Biology and Evolution*, 17(12): 1926-1932.

Arculeo, M., Brutto, S. L., Sirna-Terranova, M., Maggio, T., Cannizzaro, L. and Parrinello, N. (2003) The stock genetic structure of two Sparidae species, Diplodus vulgaris and Lithognathus mormyrus, in the Mediterranean Sea. *Fisheries Research*, 63(3): 339-347.

Arrieta, J. M., Arnaud-Haond, S. and Duarte, C. M. (2010) What lies underneath: Conserving the oceans' genetic resources. *Proceedings of the National Academy of Sciences of the United States of America* 107: 18318-18324.

Avise, J. C., Arnold, J., Ball, R. M., Bermingham, E., Lamb, T., Neigel, J. E., Reeb, C. A. and Saunders, N. C. (1987) Intraspecific Phylogeography: The Mitochondrial DNA Bridge Between Population Genetics and Systematics. *Annual Review of Ecology and Systematics* 18: 489-522.

Avise, J. C. (2000) Phylogeography: the history and formation of species. Harvard University Press.

Avise, J. C. (2009). Phylogeography: retrospect and prospect. *Journal of Biogeography* 36: 3-15.

Avise, J. C., Bowen, B. W. and Ayala, F. J. (2016) In the light of evolution X: Comparative phylogeography. *Proceedings of the National Academy of Sciences*, 113(29): 7957-7961.

Ayre, D. J., Minchinton, T. E. and Perrin, C. (2009) Does life history predict past and current connectivity for rocky intertidal invertebrates across a marine biogeographic barrier? *Molecular ecology*, 18(9):1887-1903.

Bacon, C. D., Silvestro, D., Jaramillo, C., Smith, B. T., Chakrabarty, P. and Antonelli, A. (2015) Biological evidence supports an early and complex emergence of the Isthmus of Panama. *Proceedings of the National Academy of Sciences*, 112(19): 6110-6115.

Bakun, A. (1978) Guinea current upwelling. *Nature*, 271: 147 - 150

Balloux, F. and Lugon-Moulin, N. (2002) The estimation of population differentiation with microsatellite markers. *Molecular ecology* 11: 155-65.

Bandelt, H. J., Forster, P., and Röhl, A. (1999) Median-joining networks for inferring intraspecific phylogenies. *Molecular biology and evolution*, 16 (1): 37-48.

Banks, S. C., Piggott, M. P., Williamson, J. E., Bové, U., Holbrook, N. J. and Beheregaray, L. B. (2007) Oceanic variability and coastal topography shape genetic structure in a long-dispersing sea urchin. *Ecology*, 88(12): 3055-3064.

Bargelloni, L., Alarcon, J. A., Alvarez, M. C., Penzo, E., Magoulas, A., Reis, C. and Patarnello, T. (2003) Discord in the family Sparidae (Teleostei): divergent phylogeographical patterns across the Atlantic–Mediterranean divide. *Journal of Evolutionary Biology*, 16(6): 1149-1158.

Bargelloni, L., Alarcon, J. A., Alvarez, M. C., Penzo, E., Magoulas, A., Palma, J. and Patarnello, T.(2005) The Atlantic–Mediterranean transition: discordant genetic patterns in two seabream species, Diplodus puntazzo (Cetti) and Diplodus sargus (L.). *Molecular Phylogenetics and Evolution, 36*(3): 523-535.

Barraclough, T. and Vogler, A. (2000) Detecting the Geographical Pattern of Speciation from Species-Level Phylogenies. *The American naturalist* 155: 419-434.

Bauchot, M. L. and M. M. Smith (1984) Sparidae. In Fischer, W and Bianchi, G (eds.) FAO species identification sheets for fishery purposes. Western Indian Ocean (Fishing Area 51). volume 4. FAO, Rome.

Bauchot, M.L. and Bianchi, G. (1984). *Diplodus cervinus omanensis*, Nouvele sous-espece de *Diplodus cervinus* (Lowe, 1941), capturée en mer d'Arabie (Pisces Perciformes, Sparidae). Cybium 8(103): 105

Bauchot, M. L. and J. C. Hureau (1986) Sparidae.. In P. J. P. Whitehead, M. L. Bauchot, J.-C. Hureau, J. Nielsen and E. Tortonese (eds.) Fishes of the north-eastern Atlantic and the Mediterranean. volume 2. UNESCO, Paris. pp. 883-907

Bauchot, M. L. (1987) Poissons osseux. In W. Fischer, M.L. Bauchot and M. Schneider (eds.) Fiches FAO d'identification pour les besoins de la pêche. (rev. 1). Méditerranée et mer Noire. Zone de pêche 37. Vol. II. Commission des Communautés Européennes and FAO, Rome. pp. 891-1421.

Bauchot, M. L. and Hureau, J. C. (1990) Sparidae. In Quero, J. C., Hureau, J. C., Karrer, C., Post, A., and Saldanha, L. (eds.) Check-list of the fishes of the eastern tropical Atlantic (CLOFETA). JNICT, Lisbon; SEI, Paris; and UNESCO, Paris. Vol. 2. pp. 790-812.

Beal, L. M., De Ruijter, W. P., Biastoch, A. and Zahn, R. (2011) On the role of the Agulhas system in ocean circulation and climate. *Nature*, 472(7344): 429.

Beckley, L. E. (1983). The ichthyofauna associated with *Zostera capensis* (Setchell) in the Swartkops estuary, South Africa. *South African. Journal of. Zoology.* 18: 15-24.

Beckley, L. E. (1985). The fish community of East Cape tidal pools and an assessment of nursery function of this habitat. *South African. Journal of. Zoology.* 20: 21-27.

Beckley, L. E. (1989) Larval development of Spondyliosoma emarginatimi (Cuvier & Valenciennes) (Pisces: Sparidae) from southern Africa. *South African Journal of Zoology*, 24(3): 187-192.

Bellassoued, K., Hamza, A., van Pelt, J. and Elfeki, A. (2013) Seasonal variation of Sarpa salpa fish toxicity, as related to phytoplankton consumption, accumulation of heavy metals, lipids peroxidation level in fish tissues and toxicity upon mice. *Environmental monitoring and assessment*, 185(2): 1137-1150.

Berendsen, H. J. A. (2004), De vorming van het land - Inleiding in de geologie en de geomorfologie. Fysische geografie van Nederland. Assen: Koninklijke Van Gorcum: 410

Bermingham, E., McCafferty, S. S., and Martin, A. P. (1997) Fish biogeography and molecular clocks: perspectives from the Panamanian Isthmus. *Molecular systematics of fishes*, 113-128.

Bernardi, G., Findley, L. and Rocha-Olivares, A. (2003) Vicariance and dispersal across Baja California in disjunct marine fish populations. *Evolution; international journal of organic evolution* 57: 1599-609.

Bianchi, C.N. and Morri, C. (2000) Marine biodiversity of the Mediterranean Sea: situation, problems and prospects for future research. *Marine pollution bulletin* 40(5): 367-376.

Bowen, B. W., Bass, A. L., Rocha, L. A., Grant, W. S. and Robertson, D. R. (2001) Phylogeography of the trumpetfishes (Aulostomus): ring species complex on a global scale. *Evolution*, 55: 1029–1039.

Bowen, B. W., Muss, A., Rocha, L. A. and Grant, W. S. (2006) Shallow mtDNA coalescence in Atlantic pygmy angelfishes (genus Centropyge) indicates a recent invasion from the Indian Ocean. *Journal of Heredity*, 97(1): 1-12.

Bowen, B. W., Shanker, K., Yasuda, N., Celia, M., Malay, M. C. M. D., von der Heyden, S., Paulay, G., Rocha, L. A., Selkoe, K. A., Barber, P. H. and Williams, S. T. (2014) Phylogeography unplugged: comparative surveys in the genomic era. *Bulletin of Marine Science*, 90(1): 13-46.

Bowen, B. W., Gaither, M. R., DiBattista, J. D., Iacchei, M., Andrews, K. R., Grant, W. S., Toonen, R. J. and Briggs, J. C. (2016) Comparative phylogeography of the ocean planet. *Proceedings of the National Academy of Sciences*, 113(29): 7962-7969.

Boyer, D. C., Boyer, H. J., Fossen, I. and Kreiner, A. (2001) Changes in abundance of the northern Benguela sardine stock during the decade 1990–2000, with comments on the relative importance of fishing and the environment. *African Journal of Marine Science*, 23: 67-84.

Briggs, J. C. (1974) Marine zoogeography. McGraw-Hill, New York.

Briggs, J. C. (2011) Marine extinctions and conservation. *Marine Biology* 158: 485-488.

Briggs, J.C. and Bowen, B. W. (2013) Marine shelf habitat: biogeography and evolution. *Journal of Biogeography* 40(6): 1023-1035.

Brownell, C. L. (1979). Stages in the early development of 40 marine fish species with pelagic eggs from the Cape of Good Hope. *Ichthyological Bulletin* J. L. B. Smith Institute of Ichthyology, Rhodes University 40: 1-84.

Brower, A. V. (1996) A new mimetic species of Heliconius (Lepidoptera: Nymphalidae), from southeastern Colombia, revealed by cladistic analysis of mitochondrial DNA sequences. *Zoological Journal of the Linnean Society*, *116*(3): 317-332.

Cadrin, S. X., Kerr, L. A. and Mariani, S. (2013) Stock identification methods: applications in fishery science. Academic Press, Elsevier Inc.

Canino, M. F., O'Reilly, P. T., Hauser, L. and Bentzen, P. (2005) Genetic differentiation in walleye pollock (Theragra chalcogramma) in response to selection at the pantophysin (Pan I) locus. *Canadian Journal of Fisheries and Aquatic Sciences*, 62(11): 2519-2529.

Carlsson, J., McDowell J. R., Diaz-Jaimes, P., Carlsson, J. E. L., Boles, S. B., Gold, J. R. and Graves, J. E. (2004) Microsatellite and mitochondrial DNA analyses of Atlantic bluefin tuna (*Thunnus thynnus thynnus*) population structure in the Mediterranean Sea. *Molecular Ecology* 13: 3345-3356.

Carlsson, J. (2008) Effects of microsatellite null alleles on assignment testing. *Journal of Heredity*, 99(6): 616-623.

Carlsson, J., Gauthier, D. T., Carlsson, J. E. L., Coughlan, J. P., Dillane, E., Fitzgerald, R. D., Keating, U., McGinnity, P., Mirimin, L. and Cross, T. F. (2013) Rapid, economical single- nucleotide polymorphism and microsatellite discovery based on de novo assembly of a reduced representation genome in a non- model organism: a case study of Atlantic cod Gadus morhua. *Journal of fish biology*, 82(3): 944-958.

Carstens, B. C., Pelletier, T. A., Reid, N. M. and Satler, J. D. (2013) How to fail at species delimitation. *Molecular ecology*, 22(17): 4369-4383.

Chapuis, M. P. and Estoup, A. (2007) Microsatellite null alleles and estimation of population differentiation. *Molecular biology and evolution* 24(3): 621-631.

Chevaldonne, P. (1990) Ciguatera and the saupe, Sarpa salpa (L.), in the Mediterranean: a possible misinterpretation. *Journal of fish biology*, 37(3): 503-504.

Chiba, S. N., Iwatsuki, Y., Yoshino, T. and Hanzawa, N. (2009) Comprehensive phylogeny of the family Sparidae (Perciformes: Teleostei) inferred from mitochondrial gene analyses. *Genes & genetic systems* 84: 153-70.

Chistiakov, D. A., Hellemans, B. and Volckaert, F. A. M. (2006) Microsatellites and their genomic distribution, evolution, function and applications: A review with special reference to fish genetics. *Aquaculture* 255(1-4): 1-29.

Chopelet, J., Waples, R. S. and Mariani, S. (2009) Sex change and the genetic structure of marine fish populations. *Fish and Fisheries*, 10(3): 329-343.

Claremont, M., Williams, S. T., Barraclough, T. G. and Reid, D. G. (2011) The geographic scale of speciation in a marine snail with high dispersal potential. *Journal of Biogeography* 38: 1016-1032.

Clark, P. U., Archer, D., Pollard, D., Blum, J. D., Rial, J. A., Brovkin, V., Mix, A. C., Pisias, N. G. and Roy, M. (2006) The middle Pleistocene transition:

characteristics, mechanisms, and implications for long-term changes in atmospheric pCO 2.*Quaternary Science Reviews*, 25(23): 3150-3184.

Coates, A. G. and Stallard, R. F. (2013) How old is the Isthmus of Panama? *Bulletin of Marine Science*, 89(4): 801-813.

Colborn, J., Crabtree, R. E., Shaklee, J. B., Pfeiler, E. and Bowen, B. W. (2001) The evolutionary enigma of bonefishes (Albula spp.): cryptic species and ancient separations in a globally distributed shorefish. *Evolution* 55: 807–820.

Collin, R. (2003) Phylogenetic relationships among calyptraeid gastropods and their implications for the biogeography of marine speciation. *Systematic Biology*, 52(5): 618-640.

Collins, R. A. and Cruickshank, R. H. (2013) The seven deadly sins of DNA barcoding. *Molecular ecology resources, 13*(6): 969-975.

Connell, A.D. (2012). Marine fish eggs and larvae off the east coast of South Africa. Available at: http://fisheggs-and-larvae.saiab.ac.za/

Corander, J., Majander, K. K., Cheng, L. and Merilä, J. (2013) High degree of cryptic population differentiation in the Baltic Sea herring Clupea harengus. *Molecular ecology*, 22(11): 2931-2940.

Coscia, I., Chopelet, J., Waples, R., Mann, B. and Mariani, S. (2016) Sex change and effective population size: implications for population genetic studies in marine fish. *Heredity*. 117: 251-258

Cowman, P. F. and Bellwood, D. R. (2013) Vicariance across major marine biogeographic barriers: temporal concordance and the relative intensity of hard versus soft barriers. *Proceedings of the Royal Society of London B: Biological Sciences* 280(1768): p.20131541.

Coyne, J. A. and Orr, H. A. (2004) Speciation. Sinauer Associates, Inc.

Cracraft, J. (1989). Speciation and its ontology: the empirical consequences of alternative species concepts for understanding patterns and processes of differentiation. *Speciation and its Consequences*, 28-59.

Crandall, K. A., Bininda-Emonds, O. R., Mace, G. M. and Wayne, R.K. (2000) Considering evolutionary processes in conservation biology. *Trends in ecology & evolution*, 15(7):.290-295.

Crow, J. F. and Aoki, K. (1984) Group selection for a polygenic behavioral trait: estimating the degree of population subdivision. *Proceedings of the National Academy of Sciences*, 81(19): 6073-6077.

Cruz, F., Pérez, M. and Presa, P. (2005) Distribution and abundance of microsatellites in the genome of bivalves. *Gene*, 346: 241-247.

Darriba, D., Taboada, G. L., Doallo, R. and Posada, D. (2012) jModelTest 2: more models, new heuristics and parallel computing. *Nature methods*, 9(8): 772-772.

Darwin, C. (1859) On the origin of species by means of natural selection. John Murray, .London, UK:

de Beer, C. L. (2014) Genetic and morphometric variation of Octopus vulgaris in the Benguela Current region. MSc Thesis. Rhodes University, South Africa.

DeConto, R. M. and Pollard, D. (2003) Rapid Cenozoic glaciation of Antarctica induced by declining atmospheric CO2. *Nature* 421(6920): 245-249.

de Haro, L. and Pommier, P. (2006) Hallucinatory fish poisoning (ichthyoallyeinotoxism): two case reports from the western Mediterranean and literature review. *Clinical toxicology*, 44(2): 185-188.

Degnan, J. H. and Salter, L. A. (2005) Gene tree distributions under the coalescent process. *Evolution*, 59(1): 24-37.

Demarcq, H., and Dagorne, D. (2011) Biological response associated with a coastal upwelling event. In Morales, J., Stuart, V., Platt, T. and Sathyendranath, S.(Eds). Handbook of Satellite Remote Sensing Image Interpretation: Applications for Marine Living Resources Conservation and Management, EU PRESPO and IOCCG, Dartmouth, Canada. pp217–228.

Dieckmann, U. and Doebeli, M. (1999) On the origin of species by sympatric speciation. *Nature* 400: 354-7.

Diester-Haass, L., Heine, K., Rothe, P. and Schrader, H. (1988) Late Quaternary history of continental climate and the Benguela Current off South West Africa. *Palaeogeography, palaeoclimatology, palaeoecology*, 65(1): 81-91.

Diester-Haass, L., Meyers, P. A., and Rothe, P. (1990) Miocene history of the Benguela Current and Antarctic ice volumes: Evidence from rhythmic sedimentation and current growth across the Walvis Ridge. *Paleoceanography* 5: 685–707.

Di Franco, A., Thiriet, P., Di Carlo, G., Dimitriadis, C., Francour, P., Gutiérrez, N. L., De Grissac, A. J., Koutsoubas, D., Milazzo, M., del Mar Otero, M. and Piante, C. (2016) Five key attributes can increase marine protected areas performance for small-scale fisheries management. *Scientific reports*, 6.

Di Rienzo, A., Peterson, A. C., Garza, J. C., Valdes, A. M., Slatkin, M. and Freimer, N. B. (1994) Mutational processes of simple-sequence repeat loci in human populations. Proceedings of the National Academy of Sciences, 91(8): 3166-3170.

Domingues, V. S., Bucciarelli, G., Almada, V. C. and Bernardi, G. (2005) Historical colonization and demography of the Mediterranean damselfish, Chromis chromis. *Molecular Ecology* 14(13): 4051-4063.

Duncan, K. M., Martin, A. P., Bowen, B. W. and De Couet, H. G. (2006) Global phylogeography of the scalloped hammerhead shark (*Sphyrna lewini*). *Molecular ecology* 15: 2239-51.

Duncombe Rae, C. M., Boyd, A. J. and Crawford, R. J. M. (1992) " Predation" of anchovy by an Agulhas ring: a possible contributory cause of the very poor year-class of 1989. *South African Journal of Marine Science*, *12*(1): 167-173.

Dupuis, J.R., Roe, A.D. and Sperling, F.A., 2012. Multi-locus species delimitation in closely related animals and fungi: one marker is not enough. *Molecular ecology*, 21(18):.4422-4436.

Durand, J. D., Tine, M., Panfili, J., Thiaw, O. T. and Lae, R. (2005) Impact of glaciations and geographic distance on the genetic structure of a tropical estuarine fish, Ethmalosa fimbriata (Clupeidae, S. Bowdich, 1825). *Molecular Phylogenetics and Evolution*, 36(2): 277-287.

Durand, J. D., Guinand, B., Dodson, J. J. and Lecomte, F. (2013) Pelagic Life and Depth: Coastal Physical Features in West Africa Shape the Genetic Structure of the Bonga Shad, Ethmalosa fimbriata. *PloS one*, 8(10): p.e77483.

Edgar, G. J., Russ, G. R., and Babcock, R. C. (2007) *Marine protected areas.* In: Connell, S. D., and Gillanders, B. M., (eds.) Marine Ecology. Oxford University Press, South Melbourne, VIC, Australia, pp. 533-555.

Edgar, G. J., Stuart-Smith, R. D., Willis, T. J., Kininmonth, S., Baker, S. C., Banks, S., Barrett, N. S., Becerro, M. A., Bernard, A. T., Berkhout, J. and Buxton, C. D. (2014) Global conservation outcomes depend on marine protected areas with five key features. *Nature*, 506 (7487): 216-220.

Ekman, S., (1953) Zoogeography of the sea, Sidgwick and Jackson, London, pp 417

Ellegren, H. (2004) Microsatellites: simple sequences with complex evolution. *Genetics* 5: 435-45.

Elmer, K. R. and Meyer, A. (2010) Sympatric speciation without borders? *Molecular ecology* 19: 1991-3.

El Mousadik, A. and Petit, R. J. (1996). High level of genetic differentiation for allelic richness among populations of the argan tree [Argania spinosa (L.) Skeels] endemic to Morocco. *Theoretical and Applied Genetics*, 92(7): 832-839.

Estoup, A. and Cornuet, J. M. (1999) Microsatellite evolution: inferences from population data. In: Goldstein, D. B. and Schlotterer, C. (eds) Microsatellites - Evolution and Applications. Oxford, UK: Oxford University Press. p. 49-65.

Estoup, A., Jarne, P. and Cornuet, J. M. (2002) Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. *Molecular ecology*, 11(9): 1591-1604.

Evanno, G., Regnaut, S., and Goudet, J. (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, 14: 2611–2620.

Excoffier, L. and Lischer, H. E. (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular ecology resources*, 10(3): 564-567.

FAO (2014) Fishery and Aquaculture Country Profiles: The Republic of Angola. Available at: http://www.fao.org/fishery/facp/AGO/en

Ferreri, M., Qu, W. and Han, B. (2011) Phylogenetic networks: a tool to display character conflict and demographic history. *African Journal of Biotechnology*, 10(60): 12799-12803.

Figueiredo, J., Hoorn, C., Van der Ven, P. and Soares, E. (2009) Late Miocene onset of the Amazon River and the Amazon deep-sea fan: Evidence from the Foz do Amazonas Basin. *Geology*, *37*(7): 619-622.

Floeter, S. R., Rocha, L. A., Robertson, D. R., Joyeux, J. C., Smith-Vaniz, W. F., Wirtz, P., Edwards, A. J., Barreiros, J. P., Ferreira, C. E. L., Gasparini, J. L. and Brito, A. (2008) Atlantic reef fish biogeography and evolution. *Journal of Biogeography*, 35(1): 22-47.

Florenchie, P., Lutjeharms, J. R., Reason, C. J. C., Masson, S. and Rouault, M. (2003) The source of Benguela Niños in the south Atlantic Ocean. *Geophysical Research Letters*, 30(10).

Franklin, J. (2010) Moving beyond static species distribution models in support of conservation biogeography. *Diversity and Distributions*, 16(3): 321-330.

Frisch, A. (2004) Sex-change and gonadal steroids in sequentially-hermaphroditic teleost fish. *Reviews in Fish Biology and Fisheries* 14(4): 481-499.

Fu, Y. X. (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics*, 147: 915-925.

Fujita, M. K., Leaché, A. D., Burbrink, F. T., McGuire, J. A. and Moritz, C. (2012) Coalescent-based species delimitation in an integrative taxonomy. *Trends in ecology & evolution*, 27(9): 480-488.

Funk, D. J. and Omland, K. E. (2003) Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mitochondrial DNA. *Annual Review of Ecology, Evolution, and Systematics*: 397-423.

Galarza, J. A., Carreras-Carbonell, J., Macpherson, E., Pascual, M., Roques, S., Turner, G. F. and Rico, C. (2009) The influence of oceanographic fronts and early-life-history traits on connectivity among littoral fish species. *Proceedings of the National Academy of Sciences of the United States of America* 106: 1473-8.

Galtier, N., Nabholz, B., Glémin, S. and Hurst, G. D. D. (2009) Mitochondrial DNA as a marker of molecular diversity: a reappraisal. *Molecular ecology* 18: 4541-50.

Garratt, P. A. (1985) *The Offshore Linefishery of Natal: Exploited Population Structures of the Sparids Chrysoblephus Puniceus and Cheimerius Nufar. I*. Oceanographic Research Institute.

Gibbard, P. and van Kolfschoten, T. (2004). Chapter 22: The Pleistocene and Holocene Epochs. In Gradstein, F. M.; Ogg, James G.; Smith and A. Gilbert (Eds). A Geologic Time Scale 2004. Cambridge: Cambridge University Press

Ghiselin, M. T. (1969) The evolution of hermaphroditism among animals. *Quarterly Review of Biology*: 189-208.

Gonçalves, J. M. and Erzini, K. A. R. I. M. (2000) The reproductive biology of Spondyliosoma cantharus (L.) from the SW Coast of Portugal. *Scientia Marina*, 64(4): 403-411.

Gonzalez, E. G., Beerli, P. and Zardoya, R. (2008) Genetic structuring and migration patterns of Atlantic bigeye tuna, *Thunnus obesus* (Lowe, 1839). *Bmc Evolutionary Biology*, 8.

Gopal, K., Tolley, K. A., Groeneveld, J. C. and Matthee, C. A. (2006) Mitochondrial DNA variation in spiny lobster Palinurus delagoae suggests genetically structured populations in the southwestern Indian Ocean. *Marine Ecology Progress Series*, 319(1), 191-198.

Gordon, A. L. (2003) Oceanography: The brawniest retroflection. *Nature*, 421(6926): 904-905.

Goudet, J. (1995). FSTAT (version 1.2): a computer program to calculate F-statistics. *Journal of heredity*, 86(6): 485-486.

Goudet, J., Raymond, M., de Meeüs, T., and Rousset, F. (1996). Testing differentiation in diploid populations. *Genetics*, 144(4): 1933-1940.

Grant, W. S., Leslie, R. W. and Bowen, B. W. (2005) Molecular genetic assessment of bipolarity in the anchovy genus Engraulis. *Journal of Fish Biology*, *67*(5): 1242-1265.

Grant, W. S. (2015) Problems and cautions with sequence mismatch analysis and Bayesian skyline plots to infer historical demography. *Journal of Heredity*, 106(4): 333-346.

Griffiths, M.H. and Hecht, T., 1995. On the life-history of Atractoscion aequidens, a migratory sciaenid off the east coast of southern Africa. *Journal of Fish Biology*, *47*(6): 962-985.

Guindon, S. and Gascuel, O. (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic biology*, 52(5): 696-704.

Gushchin, A. V. (2013) Feeding of fish young from littoral of the gulf Arguin (Mauritania). *Journal of Ichthyology* 53: 731-738.

Gutiérrez-García, T. A. and Vázquez-Domínguez, E. (2011) Comparative phylogeography: designing studies while surviving the process. *BioScience* 61(11): 857-868.

Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. In Nucleic acids symposium series (Vol. 41): 95-98.

Halpern, B. S. (2014) Conservation: Making marine protected areas work. *Nature*, *506* (7487): 167-168.

Hammami, I., Bahri-Sfar, L., Kaouèche, M. and Ben Hassine, O. K. (2007) Genetic characterization of striped sea bream (Lithognathus mormyrus) populations on both sides of a boundary area between eastern and western Mediterranean basins. *Cybium*, 31(2): 143-147.

Hanel, R. and Tsigenopoulos, C.S. (2011) Phylogeny, evolution, and taxonomy of sparids with some notes on their ecology and biology. In: Pavlidis, M. A., Mylonas, C.C. (eds) Sparidae: Biology and Aquaculture of Gilthead Sea Bream and Other Species. Blackwell Publishing Ltd, Oxford, UK. pp. 52–73.

Hanski, I. (1999) Metapopulation ecology. Oxford University Press. Oxford, UK.

Harpending, H. C., Batzer, M. A., Gurven, M., Jorde, L. B., Rogers, A. R. and Sherry, S. T. (1998) Genetic traces of ancient demography. *Proceedings of the National Academy of Sciences*, 95(4): 1961-1967.

Hartl, D. L. and Clark, A. G. (2007) Principles of Population Genetics. Sinauer, Sunderland, Massachusetts

Hauck, M. and Kroese, M. (2006) Fisheries compliance in South Africa: a decade of challenges and reform 1994–2004. *Marine Policy*, 30(1): 74-83.

Hauser, L. and Carvalho, G. R. (2008) Paradigm shifts in marine fisheries genetics: ugly hypotheses slain by beautiful facts. *Fish and Fisheries*, 9(4): 333-362.

Hastings, A. and Botsford, L.W. (2006) Persistence of spatial populations depends on returning home. *Proceedings of the National Academy of Sciences, 103*(15): 6067-6072.

Havelange, S., Lepoint, G., Dauby, P. and Bouquegneau, J. M. (1997) Feeding of the sparid fish Sarpa salpa in a seagrass ecosystem: diet and carbon flux. *Marine Ecology* 18(4): 289-297.

Haywood, A. M., Sellwood, B. W. and Valdes, P. J. (2000) Regional warming: Pliocene (3 Ma) paleoclimate of Europe and the Meediterranean. *Geology,* 28, 1063–1066.

Healey, A. J. E., McKeown, N. J., Potts, W. M., de Beer, C. L., Sauer, W. and Shaw, P. W. (2017) Phylogeny of the Sepia officinalis species complex in the eastern Atlantic extends the known distribution of Sepia vermiculata across the Benguela upwelling region. *African Journal of Marine Science*, 39(3): 307-313.

Hebert, P. D., Ratnasingham, S. and de Waard, J. R. (2003) Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society of London B: Biological Sciences*, 270(Suppl 1): S96-S99.

Hebert, P. D., Stoeckle, M. Y., Zemlak, T. S. and Francis, C. M. (2004) Identification of birds through DNA barcodes. *PLoS Biology*, 2(10): p.e312.

Hedrick, P. W. (1999) Perspective: highly variable loci and their interpretation in evolution and conservation. *Evolution*: 313-318.

Heemstra, P. C. and Heemstra, E. (2004). Coastal fishes of southern Africa. NISC (PTY) LTD.

Held, C. (2003) Molecular evidence for cryptic speciation within the widespread Antarctic crustacean Ceratoserolis trilobitoides (Crustacea, Isopoda). *Antarctic biology in a global context*, 135-139.

Helfrich, P. and Banner, A. H. (1960) Hallucinatory mullet poisoning: a preliminary report. *Journal of Tropical Medicine and Hygiene,* 63:.86-89.

Hellberg, M. E. (2009) Gene flow and isolation among populations of marine animals. *Annual Review of Ecology, Evolution, and Systematics,* 40: 291-310

Hemmer-Hansen, J., Therkildsen, N. O. and Pujolar, J. M. (2014) Population genomics of marine fishes: next-generation prospects and challenges. *The Biological Bulletin*, 227(2): 117-132.

Henriques, R. P. N. L. (2012). Influence of the Benguela Current in genetic sub-structuring of commercially exploited fish species. PhD thesis, Royal Holloway, University of London.

Henriques, R., Potts, W. M., Sauer, W. H. H. and Shaw, P. W. (2012) Evidence of deep genetic divergence between populations of an important recreational fishery species, Lichia amia L. 1758, around southern Africa. *African Journal of Marine Science* 34(4): 585-591.

Henriques, R., Potts, W. M., Santos, C. V., Sauer, W. H. and Shaw, P. W. (2014) Population connectivity and phylogeography of a coastal fish, Atractoscion aequidens (Sciaenidae), across the Benguela Current Region: evidence of an ancient vicariant event. *PloS one,* 9(2): p.e87907.

Henriques, R., Potts, W. M., Sauer, W. H. and Shaw, P. W. (2015) Incipient genetic isolation of a temperate migratory coastal sciaenid fish (Argyrosomus inodorus) within the Benguela Cold Current system. *Marine Biology Research* 11(4): 423-429.

Henriques, R., Potts, W. M., Sauer, W. H., Santos, C. V., Kruger, J., Thomas, J. A. and Shaw, P. W. (2016) Molecular genetic, life-history and morphological variation in a coastal warm-temperate sciaenid fish: evidence for an upwelling-driven speciation event. *Journal of Biogeography* 43(9): 1820-1831.

Henshilwood, C., d'Errico, F., Vanhaeren, M., Van Niekerk, K. and Jacobs, Z. (2004) Middle stone age shell beads from South Africa. *Science*, 304(5669): 404-404.

Hey, J. and J. Wakeley. (1997) A coalescent estimator of the population recombination rate. *Genetics* 145: 833-846.

Hey, J. and Nielsen, R. (2004) Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of Drosophila pseudoobscura and D. persimilis. *Genetics*, 167(2): 747-760.

Hey, J. and Nielsen, R. (2007) Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. *Proceedings of the National Academy of Sciences*, 104(8): 2785-2790.

Hey, J. (2009) Isolation with migration models for more than two populations. *Molecular biology and evolution*, 27(4): 905-920.

Hickerson, M. J., Meyer, C. P. and Moritz, C. (2006) DNA barcoding will often fail to discover new animal species over broad parameter space. *Systematic Biology*, 55(5): 729-739.

Hjort, A., Sheridan, S. and Davies, S. (2012) Climate change implications for fisheries of the Benguela Current region: Making the best of change? Food and Agriculture Organization of the United Nations.

Hoffmann, A. A. and Willi, Y. (2008) Detecting genetic responses to environmental change. *Nature Reviews Genetics* 9(6): 421-432.

Hohenlohe, P. A., Bassham, S., Etter, P. D., Stiffler, N., Johnson, E. A. and Cresko, W. A. (2010) Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS genetics*, 6(2): p.e1000862.

Hohenlohe, P. A., Bassham, S., Currey, M. and Cresko, W. A. (2012) Extensive linkage disequilibrium and parallel adaptive divergence across threespine stickleback genomes. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 367(1587): 395-408.

Holderegger, R., Kamm, U and Gugerli, F. (2006) Adaptive vs. neutral genetic diversity: implications for landscape genetics. *Landscape Ecology* 21(6): 797-807.

Holderegger, R. and Wagner, H. H. (2008) Landscape Genetics. *BioScience* 58: 199-208.

Holsinger, K. E. and Weir, B. S. (2009) Genetics in geographically structured populations: defining, estimating and interpreting FST. *Nature reviews. Genetics*, 10(9): 639-650.

Hsü, K. J., Ryan, W. B. F. and Cita, M. B. (1973) Late Miocene desiccation of the Mediterranean. *Nature* 242: 240-244.

Hubisz, M. J., Falush, D., Stephens, M. and Pritchard, J. K. (2009) Inferring weak population structure with the assistance of sample group information. *Molecular ecology resources*, 9(5): 1322-1332.

Hudson, R. R. and Coyne, J. A. (2002) Mathematical consequences of the genealogical species concept. *Evolution*, 56(8): 1557-1565.

Hudson, R. R. and Turelli, M. (2003) Stochasticity overrules the "three-times rule": genetic drift, genetic draft, and coalescence times for nuclear loci versus mitochondrial DNA. *Evolution*, 57(1): 182-190.

Hurt, C., Anker, A. and Knowlton, N. (2009) A multilocus test of simultaneous divergence across the Isthmus of Panama using snapping shrimp in the genus Alpheus. *Evolution: International Journal of Organic Evolution*, 63(2), 514-530.

Hutchings, L., Van der Lingen, C. D., Shannon, L. J., Crawford, R. J. M., Verheye, H. M. S., Bartholomae, C. H., Van der Plas, A. K., Louw, D., Kreiner, A., Ostrowski, M. and Fidel, Q. (2009) The Benguela Current: An ecosystem of four components. *Progress in Oceanography* 83(1): 15-32.

Iles, T. D. and Sinclair, M. (1982) Atlantic herring: stock discreteness and abundance. *Science*, 215(4533): 627-633.

Jakobsson, M., Edge, M. D. and Rosenberg, N. A.(2013) The relationship between FST and the frequency of the most frequent allele. *Genetics* 193(2): 515-528.

Jahn, B., Donner, B., Müller, P. J., Röhl, U., Schneider, R. R. and Wefer, G. (2003) Pleistocene variations in dust input and marine productivity in the northern Benguela Current: evidence of evolution of global glacial–interglacial cycles. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 193(3): 515-533.

Jombart, T. (2008) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, 24(11): 1403-1405.

Jombart, T., Devillard, S. and Balloux, F. (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC genetics*, 11(1): 94.

Joubert, C. S. W. (1981) Aspects of the biology of five species of inshore reef fishes on the Natal Coast, South Africa. *Oceanographic Research Institute Investigational Report* 51: 1-16.

Joyeux, J. C., Floeter, S. R., Ferreira, C. E. L. and Gasparini, J. L. (2001) Biogeography of tropical reef fishes: the South Atlantic puzzle. *Journal of Biogeography* 28: 831-841.

Karl, S. A., Toonen, R. J., Grant, W. S. and Bowen, B. W. (2012) Common misconceptions in molecular ecology: echoes of the modern synthesis. *Molecular Ecology* 21(17): 4171-4189.

Karl, S. A., Castro, A. L. and Garla, R. C. (2012) Population genetics of the nurse shark (Ginglymostoma cirratum) in the western Atlantic. *Marine Biology*, *159*(3): 489-498.

Kašparová, E., Van de Putte, A. P., Marshall, C. and Janko, K. (2015). Lifestyle and ice: the relationship between ecological specialization and response to Pleistocene climate change. *PloS one*, 10(11),

Kelly, D. W., MacIsaac, H. J. and Heath, D. D. (2006) Vicariance and dispersal effects on phylogeographic structure and speciation in a widespread estuarine invertebrate. *Evolution; international journal of organic evolution* 60: 257-67.

Kelly, R. P. and Palumbi, S.R. (2010) Genetic structure among 50 species of the northeastern Pacific rocky intertidal community. *PLoS One* 5(1): p.e8594.

Kensley, B. (1985) The faunal deposits of a Late Pleistocene raised beach at Milnerton, Cape Province. *Annals of South African Museum*, 95: 111–122.

Khedkar, G. D., Jamdade, R., Naik, S., David, L. and Haymer, D. (2014) DNA barcodes for the fishes of the Narmada, one of India's longest rivers. *PloS one*, 9(7): p.e101460.

Kimura, M. and Crow, J. F. (1964) The number of alleles that can be maintained in a finite population. *Genetics* 49(4): 725-738.

Kimura, M. and Ohta, T. (1978) Stepwise mutation model and distribution of allelic frequencies in a finite population. *Proceedings of the National Academy of Sciences* 75(6): 2868-2872.

Kirk, H. and Freeland, J. R. (2011) Applications and implications of neutral versus non-neutral markers in molecular ecology. *International journal of molecular sciences*, *12*(6): 3966-3988.

Knowles, L. L. and Carstens, B. C. (2007) Delimiting species without monophyletic gene trees. *Systematic biology*, 56(6): 887-895.

Kock, K. H. (1992) Antarctic fish and fisheries. Cambridge University Press.

Kraljevic, M., Dulcic, J., Cetinic, P. and Palaoro, A. (1996) Age, growth and mortality of the striped sea bream, *Lithognathus mormyrus* L., in the northern Adriatic. *Fisheries Research* 28(4): 361-370.

Krammer, R., Baumann, K., and Henrich, R. (2006) Middle to late Miocene fluctuations in the incipient Benguela Upwelling System revealed by calcareous nannofossil assemblages (ODP Site 1085A). *Palaeogeography, Palaeoclimatology, Palaeoecology*, 230: 319–334.

Landi, M., Dimech, M., Arculeo, M., Biondo, G., Martins, R., Carneiro, M., Carvalho, G. R., Brutto, S. L. and Costa, F. O. (2014) DNA barcoding for species assignment: the case of Mediterranean marine fishes. *PloS one*, 9(9): p.e106135.

Lasiak, T. A. (1981) Nursery grounds of juvenile teleosts: evidence from the surf zone of Kings Beach, Port Elizabeth. *South African Journal of Science* 77: 388-390.

Lasiak, T. A. (1986) Juveniles, food and the surf zone habitat: Implications for teleost nursery areas. *South African Journal of Zoology* 21(1): 51-56.

Lemey, P. (2009) The phylogenetic handbook: a practical approach to phylogenetic analysis and hypothesis testing. Cambridge University Press.

Lessios, H. A., Kessing, B. D. and Robertson, D. R. (1998) Massive gene flow across the world's most potent marine biogeographic barrier. *Proceedings of the Royal Society of London B: Biological Sciences*, *265*(1396): 583-588.

Lessios, H. A., Kessing, B. D., Robertson, D. R. and Paulay, G. (1999) Phylogeography of the pantropical sea urchin Eucidaris in relation to land barriers and ocean currents. *Evolution* 53: 806–817.

Lessios, H. A., Kane, J. and Robertson, D. R. (2003) Phylogeography of the pantropical sea urchin Tripneustes: contrasting patterns of population structure between oceans. *Evolution: International Journal of Organic Evolution* 57: 2026-2036.

Lessios, H. A. and Robertson, D. R. (2006) Crossing the impassable: genetic connections in 20 reef fishes across the eastern Pacific barrier. *Proceedings of the Royal Society of London B: Biological Sciences*, 273(1598): 2201-2208.

Lessios, H. A. (2008) The great American schism: divergence of marine organisms after the rise of the Central American Isthmus. *Annual Review of Ecology, Evolution, and Systematics*: 63-91.

Lett, C., Veitch, J., Van der Lingen, C. D. and Hutchings, L. (2007) Assessment of an environmental barrier to transport of ichthyoplankton from the southern to the northern Benguela ecosystems. *Marine Ecology Progress Series* 347:.247-259.

Librado, P. and Rozas, J. (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, 25 (11): 1451-1452.

Lima-Filho, P. A., de Souza Rosa, R., de Souza, A. D. S., da Costa, G. W. W. F., de Oliveira, C. and Molina, W. F. (2016) Evolutionary diversification of Western Atlantic Bathygobius species based on cytogenetic, morphologic and DNA barcode data. *Reviews in Fish Biology and Fisheries*, 26(1): 109-121.

Lisiecki, L. E. and Raymo, M. E. (2005) A Pliocene-Pleistocene stack of 57 globally distributed benthic δ18O records. *Palaeoceanography*, *20*(1): PA1003

Longhurst, A. (1998) Ecological Geography of the Sea, 1st edn. Academic Press, Cajarc, France.

Jung, G., Prange, M. and Schulz, M. (2014) Uplift of Africa as a potential cause for Neogene intensification of the Benguela upwelling system. *Nature Geoscience 7*(10): 741-747.

Lundberg, B., Ogorek, R., Galil, B. S. and Goren, M. (2004) Dietary choices of siganids fish at Shiqmona reef, Israel. *Israel Journal of Zoology* 50: 39–53.

Lutjeharms, J. R. E., and Meeuwis, J. M. (1987) The extent and variability of South-East Atlantic upwelling. *South African Journal of Marine Science* 5: 51–62.

Lutjeharms, J. R. E., Shillington, F. A., and Duncombe Rae, C. M. (1991) Observations of extreme upwelling filaments in the Southeast Atlantic Ocean. *Science*, 253: 774–776.

Lutjeharms, J. R. E. (2006) The Agulhas Current. Springer Verlang, Berlin, Heidelberg.

Lythgoe. J., and Lythgoe, G. 1971. Fishes of the sea: The coastal waters of the British Isles, Northern Europe and the Mediterranean, A photographic guide in colour. Blandford Press, London, England.

Macpherson, E. and Raventós, N. (2006) Relationship between pelagic larval duration and geographic distribution of Mediterranean littoral fishes. *Marine Ecology Progress Series*, 327: 257-265.

Maddison, W. P. and Knowles, L. L. (2006) Inferring phylogeny despite incomplete lineage sorting. *Systematic biology*, *55*(1): 21-30.

Maggs, C. A., Castilho, R., Foltz, D., Henzler, C., Jolly, M. T., Kelly, J., Olsen, J., Perez, K. E., Stam, W., Väinölä, R. and Viard, F. (2008) Evaluating signatures of glacial refugia for North Atlantic benthic marine taxa. *Ecology*, 89: 108-122

Mallet, J. (2007) Hybrid speciation. *Nature* 446: 279-83.

Mann, B. Q. (1992) Aspects of the biology of two inshore sparid fishes (Diplodus sargus capensis and Diplodus cervinus hottentotus) off the south-east coast of South Africa. MSc thesis, Rhodes University, South Africa.

Mann, B. Q. and Buxton, C. D. (1997). Age and growth of *Diplodus sargus capensis* and *D. cervinus hottentotus* (Sparidae) on the Tsitsiskamma coast, South Africa. *Cybium* 21: 135-147.

Mann, B. Q. and Buxton, C. D. (1998) The reproductive biology of *Diplodus sargus capensis* and *D. cervinus hottentotus* (Sparidae) off the south-east Cape coast, South Africa. *Cybium*, 22: 31-47.

Mann, B. Q. and Dunlop, S. W. (2013) Sarpa salpa. In: B.Q. Mann (ed.), Southern African Marine Linefish Species Profiles, Special Publication No. 9 Oceanographic Research Institute, Durban. pp. 273-274.

Mann, B. Q., Buxton, C. D., Russell, B., Pollard, D. and Carpenter, K. E. (2014). *Spondyliosoma emarginatum. The IUCN Red List of Threatened Species 2014*: e.T170172A1287155.

Marko, P. B., Hoffman, J. M., Emme, S. A., McGovern, T. M., Keever, C. C. and Nicole Cox, L. (2010) The 'Expansion–Contraction'model of Pleistocene biogeography: rocky shores suffer a sea change? *Molecular Ecology*, 19(1): 146-169.

Marko, P. B., Eytan, R. I. and Knowlton, N. (2015) Do large molecular sequence divergences imply an early closure of the Isthmus of Panama? *Proceedings of the National Academy of Sciences* 112(43): E5766-E5766.

Marlow, J. R., Lange, C. B., Wefer, G., and Rosell-Melé, A. (2000) Upwelling intensification as part of the Pliocene-Pleistocene climate transition. *Science* 290: 2288–2291.

Matthee, C. A, Fourie, F., Oosthuizen, W. H., Meyër, M. A. and Tolley, K. A. (2006) Mitochondrial DNA sequence data of the Cape fur seal (Arctocephalus pusillus pusillus) suggest that population numbers may be affected by climatic shifts. *Marine Biology* 148: 899-905.

Matthee, C. A, Cockcroft, A. C., Gopal, K. and Von Der Heyden, S. (2007) Mitochondrial DNA variation of the west-coast rock lobster, Jasus lalandii: marked genetic diversity differences among sampling sites. *Marine and Freshwater Research* 58: 1130-1135.

Matz, M.V. and Nielsen, R. (2005) A likelihood ratio test for species membership based on DNA sequence data. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *360*(1462):1969-1974.

Mayr, E. (1954) Geographic speciation in tropical echinoids. *Evolution*: 1-18.

McCartney, M. A., Keller, G. and Lessios, H. A. (2000) Dispersal barriers in tropical oceans and speciation in Atlantic and eastern Pacific sea urchins of the genus Echinometra. *Molecular Ecology* 9: 1391-1400.

McKeown, N. J., Arkhipkin, A. I. and Shaw, P. W. (2017) Regional genetic population structure and fine scale genetic cohesion in the Southern blue whiting Micromesistius australis. *Fisheries Research, 185*: 176-184.

McCook, L. J., Almany, G. R., Berumen, M. L., Day, J. C., Green, A. L., Jones, G. P., Leis, J. M., Planes, S., Russ, G. R., Sale, P. F. and Thorrold, S. R. (2009) Management under uncertainty: guide-lines for incorporating connectivity into the protection of coral reefs. *Coral Reefs*, 28(2): 353-366.

McCook, L. J., Ayling, T., Cappo, M., Choat, J. H., Evans, R. D., De Freitas, D. M., Heupel, M., Hughes, T. P., Jones, G. P., Mapstone, B. and Marsh, H. (2010) Adaptive management of the Great Barrier Reef: a globally significant demonstration of the benefits of networks of marine reserves. *Proceedings of the National Academy of Sciences*, 107(43): 18278-18285.

Meglecz, E., Petenian, F., Danchin, E., D'Acier, A. C., Rasplus, J. Y. and Faure, E (2004) High similarity between flanking regions of different microsatellites detected within each of two species of Lepidoptera: Parnassius apollo and Euphydryas aurinia. *Molecular Ecology*, 13(6): 1693-1700.

Meier, R. (2008) DNA sequences in taxonomy: opportunities and challenges. In: Wheeler, Q. D.(Ed), The New Taxonomy. CRC Press, Taylor and Francis Group, pp: 95-127.

Meier, R., Zhang, G. and Ali, F. (2008) The use of mean instead of smallest interspecific distances exaggerates the size of the "barcoding gap" and leads to misidentification. *Systematic biology*, 57(5): 809-813.

Meyer, A., Morrissey, J. M. and Schartl, M. (1994) Recurrent origin of a sexually selected trait in Xiphophorus fishes inferred from a molecular phylogeny. *Nature*, 368(6471): 539-542.

Milano, I., Babbucci, M., Cariani, A., Atanassova, M., Bekkevold, D., Carvalho, G. R., Espiñeira, M., Fiorentino, F., Garofalo, G., Geffen, A. J. and Hansen, J. (2014) Outlier SNP markers reveal fine- scale genetic structuring across European hake populations (Merluccius merluccius). *Molecular Ecology*, 23(1): 118-135.

Miller, P. J., and Loates, M. J. (1997) Collins Pocket Guide: Fish of Britain and Europe. Harper Collins Publishers, London, England.

Miralles, L., Juanes, F., Pardiñas, A. F. and Garcia-Vazquez, E. (2014) Paleoclimate shaped bluefish structure in the northern hemisphere. *Fisheries*, 39(12): 578-586.

Monaghan, M. T., Wild, R., Elliot, M., Fujisawa, T., Balke, M., Inward, D. J., Lees, D. C., Ranaivosolo, R., Eggleton, P., Barraclough, T. G. and Vogler, A. P. (2009) Accelerated species inventory on Madagascar using coalescent-based models of species delineation. *Systematic biology*: p.syp027.

Moran, P. and Kornfield, I. (1993) Retention of an ancestral polymorphism in the Mbuna species flock (Teleostei: Cichlidae) of Lake Malawi. *Molecular Biology and Evolution*, 10: 1015-1027.

Moritz, C. (1994) Defining 'evolutionarily significant units' for conservation. *Trends in ecology and evolution*, 9(10): 373-374.

Moritz, C. and Cicero, C. (2004) DNA barcoding: promise and pitfalls. *PLoS Biol*, *2*(10): p.e354.

Mouine, N., Ktari, M. H. and Chakroun-Marzouk, N. (2011) Reproductive characteristics of Spondyliosoma cantharus (Linnaeus, 1758) in the Gulf of Tunis. *Journal of Applied Ichthyology* 27(3): 827-831.

Moura, R. L., Amado-Filho, G. M., Moraes, F. C., Brasileiro, P. S., Salomon, P. S., Mahiques, M. M., Bastos, A. C., Almeida, M. G., Silva, J. M., Araujo, B. F. and Brito, F. P. (2016) An extensive reef system at the Amazon River mouth. *Science advances*, 2(4): p.e1501252.

Muss, A., Ross Robertson, D., Stepien, C. A., Wirtz, P. and Bowen, B. W. (2001) Phylogeography of Ophioblennius: the role of ocean currents and geography in reef fish evolution. *Evolution*, *55*(3): 561-572.

Neethling, M., Matthee, C. A, Bowie, R. C. K. and Von der Heyden, S. (2008) Evidence for panmixia despite barriers to gene flow in the southern African endemic, Caffrogobius caffer (Teleostei: Gobiidae). *BMC evolutionary biology* 8: 325.

Nielsen, R. and Wakeley, J. (2001) Distinguishing migration from isolation: a Markov chain Monte Carlo approach. *Genetics*, 158(2): 885-896.

Nikaido, M., Sasaki, T., Emerson, J. J., Aibara, M., Mzighani, S. I., Budeba, Y. L., Ngatunga, B. P., Iwata, M., Abe, Y., Li, W.-H. and Okada, N. (2011) Genetically distinct coelacanth population off the northern Tanzanian coast. *Proceedings of the National Academy of Sciences*: 1-5.

Nosil, P. and Schluter, D. (2011) The genes underlying the process of speciation. *Trends in Ecology & Evolution*, 26(4): 160-167.

Nwani, C. D., Becker, S., Braid, H. E., Ude, E. F., Okogwu, O. I. and Hanner, R. (2011) DNA barcoding discriminates freshwater fishes from southeastern Nigeria and provides river system-level phylogeographic resolution within some species. *Mitochondrial DNA*, *22*(sup1): 43-51.

Olivar, P. M. and Shelton, P. A. (1993) Larval Fish Assemblages of the Benguela Current. *Bulletin of Marine Science* 53: 450-474.

Olsen, J. L., Stam, W. T., Coyer, J. A., Reusch, T. B., Billingham, M., Boström, C., Calvert, E., Christie, H., Granger, S., Lumiere, R. L. and Milchakova, N. (2004) North Atlantic phylogeography and large-scale population differentiation of the seagrass Zostera marina L *Molecular ecology* 13(7): 1923-1941.

Olson, D. M., Dinerstein, E., Wikramanayake, E. D., Burgess, N. D., Powell, G. V., Underwood, E. C., D'amico, J. A., Itoua, I., Strand, H. E., Morrison, J. C. and Loucks, C. J. (2001) Terrestrial Ecoregions of the World: A New Map of Life on Earth A new global map of terrestrial ecoregions provides an innovative tool for conserving biodiversity. *BioScience*, 51(11): 933-938

Ovenden, J. R., Kashiwagi, T., Broderick, D., Giles, J. and Salini, J. (2009) The extent of population genetic subdivision differs among four co-distributed shark species in the Indo-Australian archipelago. *BMC evolutionary biology*, *9*(1), p.1.

Paetkau, D., Slade, R., Burden, M. and Estoup, A. (2004) Genetic assignment methods for the direct, real- time estimation of migration rate: a simulation- based exploration of accuracy and power. *Molecular ecology*, 13(1): 55-65.

Pajuelo, J. G. and Lorenzo, J. M. (1999) Life history of black seabream, Spondyliosoma cantharus, off the Canary Islands, Central-east Atlantic. *Environmental Biology of fishes*, 54(3): 325-336.

Pajuelo, J. G. and Lorenzo, J. M. (2001). Biology of the annular seabream, *Diplodus annularis* (Sparidae), in coastal waters of the Canary Islands. *Journal of Applied Ichthyology* 17: 121–125.

Pajuelo, J., Lorenzo, J., Domínguez, R., Ramos, A. and Gregoire, M. (2003)a. On the Population Ecology of the Zebra Seabream Diplodus cervinus cervinus (Lowe 1838) from the Coasts of the Canarian archipelago, North West Africa. *Environmental Biology of fishes* 67(4): 407-416.

Pajuelo, J. G., Lorenzo, M. J. and Dominguez-Seoane, R. (2003)b. Age estimation and growth of the zebra seabream *Diplodus cervinus cervinus* (Lowe 1838) on the Canary Islands shelf (Central East Atlantic). *Fisheries Research* 62 (1) 97 –103.

Palsbøll, P. J., Berube, M. and Allendorf, F. W.(2007) Identification of management units using population genetic data. *Trends in Ecology & Evolution*, 22(1): 11-16.

Palumbi, S. R. (1992). Marine speciation on a small planet. *Trends in Ecology & Evolution* 7: 114-118.

Palumbi, S (1996) Nucleic acids II: the polymerase chain reaction. In: Molecular systematics, 2nd edn (eds Hills, D. M., Moritz, C. and Mable, B. K.). Sinauer Associates, Sunderland.

Palumbi, S. R., Grabowsky, G., Duda, T., Geyer, L. and Tachino, N. (1997) Speciation and population genetic structure in tropical Pacific sea urchins. *Evolution* 51: 1506-1517.

Palumbi, S. R. (2004) Marine Reserves and Ocean Neighborhoods: The Spatial Scale of Marine Populations and Their Management. *Annual Review of Environment and Resources 29*(1): 31-68.

Panchal, M. and Beaumont, M. A. (2007) The automation and evaluation of nested clade phylogeographic analysis. *Evolution*, 61(6): 1466-1480.

Papadopoulou, A. and Knowles, L. L. (2016) Toward a paradigm shift in comparative phylogeography driven by trait-based hypotheses. *Proceedings of the National Academy of Sciences*, 113(29): 8018-8024.

Parmesan, C., Gaines, S., Gonzalez, L., Kaufman, D. M., Kingsolver, J., Townsend Peterson, A. and Sagarin, R. (2005) Empirical perspectives on species borders: from traditional biogeography to global change.*Oikos*,108(1): 58-75.

Patarnello, T., Volckaert, F.A. and Castilho, R. (2007) Pillars of Hercules: is the Atlantic–Mediterranean transition a phylogeographical break? *Molecular ecology* 16(21): 4426-4444.

Pawson, M., G. (1995) Biogeographical identification of English Channel fish and shellfish stocks. Fisheries Research Technical Report (number 99), MAFF

Direct Fisheries Research Lowestoft, England. Available at: https://www.cefas.co.uk/publications/techrep/tech99.pdf

Peeters, F. J. C., Acheson, R., Brummer, G. J. A., de Ruijter, W. P. M., Schneider, R. R., Ganssen, G. M., Ufkes, E. and Kroon, D. (2004) Vigorous exchange between the Indian and Atlantic Oceans at the end of the past five glacial periods. *Nature* 430: 661–665.

Penven, P., Lutjeharms, J. R. E., Marchesiello, P., Roy, C. and Weeks, S.J. (2001) Generation of cyclonic eddies by the Agulhas Current in the lee of the Agulhas Bank. *Geophysical Research Letters* 28: 1055–1058.

Pereira, L.H., Hanner, R., Foresti, F. and Oliveira, C., 2013. Can DNA barcoding accurately discriminate megadiverse Neotropical freshwater fish fauna? *BMC genetics*, *14*(1): 1.

Perez-Losada, M., Nolte, M. J., Crandall, K. A. and Shaw, P. W. (2007) Testing hypotheses of population structuring in the Northeast Atlantic Ocean and Mediterranean Sea using the common cuttlefish Sepia officinalis. *Molecular Ecology* 16: 2667-2679.

Perez, L., Infante, C., Ponce, M., Crespo, A., Zuasti, E., Funes, V., Catanese, G. and Manchado, M. (2008) Characterization of eight microsatellite markers in the white sea bream, Diplodus sargus (Teleostei, Sparidae). *Molecular ecology resources*, 8(6): 1291-1293.

Perez-Ruzafa, A., Gonzalez-Wanguemert, M., Lenfant, P., Marcos, C. and Garcia-Charton, J. A. (2006) Effects of fishing protection on the genetic structure of fish populations. *Biological Conservation*, 129: 244-255.

Peterson, R. G. and Stramma, L. (1991) Upper-level circulation in the South Atlantic Ocean. *Progress in oceanography*, 26(1): 1-73.

Petrick, B., McClymont, E. L., Felder, S., Rueda, G., Leng, M. J. and Rosell-Melé, A. (2015) Late Pliocene upwelling in the Southern Benguela region. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 429: 62-71.

Petrick, B. F., McClymont, E. L., Marret, F. and Meer, M. T. (2015) Changing surface water conditions for the last 500 ka in the Southeast Atlantic: Implications for variable influences of Agulhas leakage and Benguela upwelling. *Paleoceanography*, 30(9): 1153-1167.

Pirazzoli, P. A. (1997) Sea-level changes: the last 20 000 years. *Oceanographic Literature Review* 8 (44): 785.

Piry, S., Alapetite, A., Cornuet, J. M., Paetkau, D., Baudouin, L. and Estoup, A. (2004) GENECLASS2: a software for genetic assignment and first-generation migrant detection. *Journal of heredity*, 95(6): 536-539.

Posada, D. and Crandall, K. A. (2001) Intraspecific gene genealogies: trees grafting into networks. *Trends in ecology & evolution*, 16(1): 37-45.

Posada, D. and Buckley, T. R. (2004) Model selection and model averaging in phylogenetics: advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. *Systematic biology*, 53(5): 793-808.

Posada, D. (2009) Selection of Models of DNA Evolution with jModelTest. *Bioinformatics for DNA sequence analysis*, pp.93-112.

Potts, W. M., Henriques, R., Santos, C. V., Munnik, K., Ansorge, I., Dufois, F., Booth, A. J., Kirchner, C., Sauer, W. H. and Shaw, P. W. (2014) Ocean warming, a rapid distributional shift, and the hybridization of a coastal fish species. *Global change biology* 20(9): 2765-2777.

Potts, W. M., Childs, A. R., Sauer, W. H. H. and Duarte, A. D. C. (2009) Characteristics and economic contribution of a developing recreational fishery in southern Angola. *Fisheries management and ecology*, 16(1): 14-20.

Potts, W. M., Gotz, A., and James, N. (2015) Review of the projected impacts of climate change on coastal fishes in southern Africa. *Reviews in Fish Biology and Fisheries* 25: 603-630.

Pritchard, J. K., Stephens, M. and Donnelly, P. (2000) Inference of population structure using multilocus genotype data. *Genetics*, *155*(2): 945-959.

Purcell, J.F., Cowen, R.K., Hughes, C.R. and Williams, D.A. (2006) Weak genetic structure indicates strong dispersal limits: a tale of two coral reef fish. *Proceedings of the Royal Society of London B: Biological Sciences* 273(1593): 1483-1490.

Putman, A. I. and Carbone, I. (2014) Challenges in analysis and interpretation of microsatellite data for population genetic studies. *Ecology and evolution*, 4(22): 4399-4428.

Quenouille, B., Hubert, N., Bermingham, E. and Planes, S. (2011) Speciation in tropical seas: allopatry followed by range change. *Molecular phylogenetics and evolution* 58(3):.546-552.

R Core Team. (2017). R: A language and environment for statistical computing. Available at: http://www.r-project.org/

Rambaut, A. (2012) FigTree v. 1.4.0. Available at http://tree.bio.ed.ac.uk/software/figtree

Ramsay, P. J. and Cooper, J. A. G. (2002) Late Quaternary sea-level change in South Africa. *Quaternary Research*, 57(1): 82-90.

Ranzi, S. (1933) Sparidae. In: Uova, larve e stadi giovanili di Teleostei. *Fauna e Flora del Golfo di Napoli Monograph* 38: 330-375.

Raymond, M. and Rousset, F (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity*, 86: 248-249

Richardson, T. J. (2010) The taxonomy, life-history and population dynamics of blacktail, *Diplodus capensis* (Perciformes: Sparidae), in southern Angola. In: *Department of Ichthyology and Fishery Sciences*, Rhodes University, Grahamstown.

Richardson, T. J., Potts, W. M., Santos, C. V. and Sauer, W. H. (2011) Comparison of the population structure and life-history parameters of Diplodus capensis (Sparidae) in exploited and unexploited areas of southern Angola. *African Journal of Marine Science*, 33(2): 191-201.

Richter, I., Behera, S. K., Masumoto, Y., Taguchi, B., Komori, N. and Yamagata, T. (2010) On the triggering of Benguela Niños: Remote equatorial versus local influences. *Geophysical Research Letters*, 37(20).

Riginos, C. and Victor, B.C. (2001) Larval spatial distributions and other early life–history characteristics predict genetic differentiation in eastern Pacific blennioid

fishes. *Proceedings of the Royal Society of London B: Biological Sciences* 268(1479): 1931-1936.

Rissler, L. J. (2016) Union of phylogeography and landscape genetics. *Proceedings of the National Academy of Sciences 113*(29): 8079-8086.

Rocha, L. A., Bass, A. L., Robertson, D. R. and Bowen, B. W. (2002) Adult habitat preferences, larval dispersal, and the comparative phylogeography of three Atlantic surgeonfishes (Teleostei: Acanthuridae). *Molecular Ecology*, 11(2): 243-251.

Rocha, L. A., Robertson, D., Rocha, C. R., Tassell, J. L., Craig, M. T. and Bowen, B. W. (2005) Recent invasion of the tropical Atlantic by an Indo-Pacific coral reef fish. *Molecular Ecology*, 14(13): 3921-3928.

Rogers, A. R. and Harpending, H. (1992) Population growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology and Evolution*. 9: 552-569.

Rogers, A. R. (1995) Genetic evidence for a Pleistocene population explosion. *Evolution*, 49(4): 608-615.

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M. A. and Huelsenbeck, J. P. (2012) MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic biology*, 61(3): 539-542.

Roques, S., Galarza, J. A., Macpherson, E., Turner, G. F. and Rico, C. (2007a) Isolation and characterization of nine polymorphic microsatellite markers in the two-banded sea bream (Diplodus vulgaris) and cross-species amplification in the white sea bream (Diplodus sargus) and the saddled bream (Oblada melanura). *Molecular Ecology Notes*, 7(4): 661-663

Roques, S., Galarza, J. A., Macpherson, E., Turner, G. F., Carreras-Carbonell, J. and Rico, C. (2007b) Isolation of eight microsatellites loci from the saddled bream, Oblada melanura and cross-species amplification in two sea bream species of the genus Diplodus. *Conservation Genetics*, 8(5): 1255-1257.

Roux, C., Fraisse, C., Romiguier, J., Anciaux, Y., Galtier, N. and Bierne, N. (2016) Shedding light on the grey zone of speciation along a continuum of genomic divergence. *PLoS biology*, 14(12): p.e2000234.

Roy, K., Jablonsk, D. and Valentine, J. W. (1995) Thermally anomalous assemblages revisited: patterns in the extraprovincial latitudinal range shifts of Pleistocene marine mollusks. *Geology*, 23(12): 1071-1074.

Rozen, S. and Skaletsky, H. (1999) Primer3 on the WWW for general users and for biologist programmers. *Bioinformatics methods and protocols*, 365-386.

Rundle, H. D. and Nosil, P. (2005) Ecological speciation. *Ecology Letters* 8: 336-352.

Russell, F. S. (1976) The eggs and planktonic stages of British marine fishes. Academic Press, London. pp 524

Russell, B., Carpenter, K. E., Pollard, D., Mann, B. Q. and Buxton, C. D. (2014a) *Lithognathus mormyrus. The IUCN Red List of Threatened Species 2014*: e.T170160A1284573.

Russell, B., Pollard, D., Mann, B. Q., Buxton, C. D. and Carpenter, K. E. (2014b) *Sarpa salpa*. *The IUCN Red List of Threatened Species 2014*: e.T170169A1286510.

Russell, B., Pollard, D. & Carpenter, K. E. (2014c) *Spondyliosoma cantharus*. *The IUCN Red List of Threatened Species:* e.T170258A1303321.

Ruzzante, D. E., Mariani, S., Bekkevold, D., André, C., Mosegaard, H., Clausen, L. A., Dahlgren, T. G., Hutchinson, W. F., Hatfield, E. M., Torstensen, E. and Brigham, J. (2006) Biocomplexity in a highly migratory pelagic marine fish, Atlantic herring. *Proceedings of the Royal Society of London B: Biological Sciences*, 273(1593): 1459-1464.

Ryman, N., Utter, F. and Laikre, L. (1995) Protection of intraspecific biodiversity of exploited fishes. *Reviews in Fish Biology and Fisheries*, 5(4): 417-446.

Ryman, N. and Palm, S. (2006) POWSIM: a computer program for assessing statistical power when testing for genetic differentiation. *Molecular Ecology Resources*, 6(3): 600-602.

SADSAA. (2012) South African Deep Sea Angling Association. Available at: http://www.sadsaa.com/

Sala-Bozano, M., Tsalavouta, M. and Mariani, S. (2009) Isolation and characterisation of new polymorphic microsatellite markers for the striped sea bream (Lithognathus mormyrus). *Conservation Genetics*, 10(5): 1507-1509.

Sala-Bozano, M., Ketmaier, V. and Mariani, S. (2009) Contrasting signals from multiple markers illuminate population connectivity in a marine fish. *Molecular ecology* 18: 4811-26.

Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Press, Cold Spring Harbor, New York

Santini, F., Carnevale, G. and Sorenson, L. (2014) First multi-locus timetree of seabreams and porgies (Percomorpha: Sparidae). *Italian Journal of Zoology*, *81*(1): 55-71.

Savolainen, V., Anstett, M. C., Lexer, C., Hutton, I., Clarkson, J. J., Norup, M. V., Powell, M. P., Springate, D., Salamin, N. and Baker, W. J. (2006) Sympatric speciation in palms on an oceanic island. *Nature* 441: 210-3.

Schlick-Steiner, B. C., Steiner, F. M., Seifert, B., Stauffer, C., Christian, E. and Crozier, R. H. (2010) Integrative taxonomy: a multisource approach to exploring biodiversity. *Annual review of entomology*, 55: 421-438.

Schluter, D. (2001) Ecology and the origin of species. *Trends in ecology & evolution* 16(7): 372-380.

Schluter, D. and Conte, G. L. (2009). Genetics and ecological speciation. *Proceedings of the National Academy of Sciences of the United States of America* 106: 9955-62.

Schneider, S. and Excoffier, L. (1999). Estimation of past demographic parameters from the distribution of pairwise differences when the mutation rates vary among sites: application to human mitochondrial DNA. *Genetics*, 152(3):1079-1089.

Selkoe, K. A. and Toonen, R. J. (2006) Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. *Ecology letters* 9(5): 615-629.

Selkoe, K. A., Henzler, C. M. and Gaines, S. D. (2008) Seascape genetics and the spatial ecology of marine populations. *Fish and Fisheries* 9(4): 363-377.

Shannon, L. V, Walters, N. M., and Moldan, A. G. S. (1983) Some features in two Cape bays as deduced from satellite ocean-colour imagery. *South African Journal of Marine Science* 1: 111–122.

Shannon, L. V. (1985) The Benguela Ecosystem .1. Evolution of the Benguela, Physical Features and Processes. *Oceanography and Marine Biology* 23:105- 182

Shannon, L. V, and Nelson, G. (1996) The Benguela: Large scale features and processes and system variability. In: G. Wefer, G., Berger, W. H., Siedler, G., and Webb. D. J.(Eds) The South Atlantic: Present and Past Circulation. Springer, Berlin. pp. 163–210.

Shannon, L. V. and O'Toole, M. J. (2003) Sustainability of the Benguela: ex Africa semper aliquid novi. In Hempel, G. and Sherman, K.(eds) Large Marine Ecosystems of the World. Trends in Exploitation Protection and Research, The Benguela Current. Elsevier Verlag. pp 227-253.

Shen, K. N., Jamandre, B. W., Hsu, C. C., Tzeng, W. N. and Durand, J. D., 2011. Plio-Pleistocene sea level and temperature fluctuations in the northwestern Pacific promoted speciation in the globally-distributed flathead mullet Mugil cephalus. *BMC Evolutionary Biology*, 11(1): p.1.

Sherman, K. and Hempel, G. (Eds) (2008) The UNEP Large Marine Ecosystem Report: A perspective on changing conditions in LMEs of the world's Regional Seas. UNEP Regional Seas Report and Studies No. 182. United Nations Environment Programme. Nairobi, Kenya.

Shillington F. A, Duncombe, R., Florenchie and Penven. (2004) Large scale physical variability of the Benguela Current Large Marine Ecosystem (BCLME). In: Shannon, G. H., Malanotte-Rizzoli. P., Moloney, C. and Woods, J. (Eds) Benguela: Predicting a Large Marine Ecosystem. Elsevier, Amsterdam, The Netherlands.

Shulman, M. J. and Bermingham, E. (1995) Early life histories, ocean currents, and the population genetics of Caribbean reef fishes. *Evolution* 49: 897–910.

Siesser, W. G. (1980) Late Miocene origin of the Benguela upwelling system off northern Namibia. *Science*, 208: 283–285.

Silva, G., Horne, J. B. and Castilho, R. (2014) Anchovies go north and west without losing diversity: post-glacial range expansions in a small pelagic fish. *Journal of biogeography*, *41*(6): 1171-1182

Sites Jr, J. W. and Marshall, J. C. (2004) Operational criteria for delimiting species. *Annual Review of Ecology, Evolution, and Systematics*: 199-227.

Slatkin, M. (1994). Gene flow and population structure. In: Real, L. A (ed.) Ecological Genetics, Princeton University Press, Princeton, NJ. pp. 3–17.

Slatkin, M. (1995) A measure of population subdivision based on microsatellite allele frequencies. *Genetics*, 139(1): 457-462.

Smith, J. L. B. and Smith, M. M. (1986) Sparidae. In: M.M. Smith, M. M. and Heemstra, P. C. (eds.) Smiths' sea fishes. Springer-Verlag, Berlin. pp. 580-594

Soekoe, M. (2016) Adaptations in allopatric populations of Triakis megalopterus isolated by the Benguela Current. Steps towards understanding evolutionary processes affecting regional biodiversity. PhD Thesis. Rhodes University, South Africa.

Sousa, V. C., Grelaud, A. and Hey, J. (2011) On the nonidentifiability of migration time estimates in isolation with migration models. *Molecular ecology*, 20(19):.3956-3962.

Southern Science (1995) A study of the black bream spawning ground at Littlehampton. Report No: 95/2/1147. (Report to: United Marine Dredging, ARC Marine Ltd. and South Coast Shipping). Southern Science Ltd. Hants. England.

Spalding, M. D., Fox, H. E., Allen, G. R., Davidson, N., Ferdaña, Z. A., Finlayson, M. A. X., Halpern, B. S., Jorge, M. A., Lombana, A. L., Lourie, S. A. and Martin, K. D. (2007) Marine ecoregions of the world: a bioregionalization of coastal and shelf areas. *BioScience*, 57(7): 573-583.

Storfer, A., Murphy, M. A, Spear, S. F., Holderegger, R. and Waits, L. P. (2010) Landscape genetics: where are we now? *Molecular ecology* 19: 3496-514.

Summerer, M., Hanel, R. and Sturmbauer, C. (2001) Mitochondrial phylogeny and biogeographic affinities of sea breams of the genus Diplodus (Sparidae). *Journal of Fish Biology*, 59(6): 1638-1652.

Sunday, J. M., Bates, A. E. and Dulvy, N. K. (2012) Thermal tolerance and the global redistribution of animals. *Nature Climate Change*, 2(9): 686-690.

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123: 585-595.

Takahata, N. and Nei, M. (1985) Gene genealogy and variance of interpopulational nucleotide differences. *Genetics*, 110(2): 325-344.

Tamura, K., Stecher, G., Peterson, D., Filipski, A. and Kumar, S. (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. *Molecular biology and evolution*, 30(12): 2725-2729.

Taylor, M. S. and Hellberg, M. E. (2005). Marine radiations at small geographic scales: speciation in neotropical reef gobies (Elacatinus). *Evolution: International Journal of Organic Evolution* 59: 374-385.

Teske, P. R., Oosthuizen, A., Papadopoulos, I. and Barker, N. P. (2007) Phylogeographic structure of Octopus vulgaris in South Africa revisited: identification of a second lineage near Durban harbour. *Marine Biology*, *151*(6): 2119-2122.

Teske, P. R., Von Der Heyden, S., McQuaid, C. D. and Barker, N. P. (2011). A review of marine phylogeography in southern Africa. *South African Journal of Science* 107: 1-11.

Teske, P. R., Zardi, G. I., McQuaid, C., and Nicastro, K. R. (2013) Two sides of the same coin: Extinctions and originations across the Atlantic/Indian Ocean boundary as consequences of the same climate oscillation. *Frontiers of Biogeography* 5: 217–220.

Therkildsen, N. O., Hemmer-Hansen, J., Als, T. D., Swain, D.P ., Morgan, M. J., Trippel, E. A., Palumbi, S. R., Meldrup, D. and Nielsen, E. E. (2013)

Microevolution in time and space: SNP analysis of historical DNA reveals dynamic signatures of selection in Atlantic cod. *Molecular ecology*, 22(9): 2424-2440.

Thompson, K.F., Patel, S., Baker, C.S., Constantine, R. and Millar, C.D. (2016) Bucking the trend: genetic analysis reveals high diversity, large population size and low differentiation in a deep ocean cetacean. *Heredity* 116: 277-285

Tine, M., Kuhl, H., Gagnaire, P. A., Louro, B., Desmarais, E., Martins, R. S., Hecht, J., Knaust, F., Belkhir, K., Klages, S. and Dieterich, R. (2014) European sea bass genome and its variation provide insights into adaptation to euryhalinity and speciation. *Nature communications*, 5: p.5770.

Valdes, A. M., Slatkin, M. and Freimer, N. B. (1993) Allele frequencies at microsatellite loci: the stepwise mutation model revisited. *Genetics* 133(3): 737-749

Van der Walt, B. A. and Mann, B. Q. (1998) Aspects of the reproductive biology of *Sarpa salpa* off the east coast of South Africa. *South African Journal of Marine Science* 33(4): 241-248.

Veitch, J. A., Florenchie, P., and Shillington, F. A. (2006) Seasonal and interannual fluctuations of the Angola–Benguela Frontal Zone (ABFZ) using 4.5 km resolution satellite imagery from 1982 to 1999. *International Journal of Remote Sensing* 27: 987–998.

Veríssimo, A., McDowell, J. R. and Graves, J. E. (2010) Global population structure of the spiny dogfish Squalus acanthias, a temperate shark with an antitropical distribution. *Molecular Ecology*, 19(8): 1651-1662.

Villamil, M. M., Lorenzo, J. M., Pajuelo, J. G., Ramos, A. and Coca, J. (2002) Aspects of the life history of the salema, *Sarpa salpa* (Pisces, Sparidae), off the Canarian Archipelago (central-east Atlantic). *Environmental Biology of Fishes* 63: 183-192.

Von der Heyden, S., Lipinski, M. R. and Matthee, C. A. (2007) Mitochondrial DNA analyses of the Cape hakes reveal an expanding, panmictic population for Merluccius capensis and population structuring for mature fish in Merluccius paradoxus. *Molecular Phylogenetics and Evolution* 42: 517-527.

Von der Heyden, S., Prochazka, K. and Bowie, R. C. K. (2008) Significant population structure and asymmetric gene flow patterns amidst expanding populations of Clinus cottoides (Perciformes, Clinidae): application of molecular data to marine conservation planning in South Africa. *Molecular ecology* 17: 4812-26.

Von der Heyden, S., Lipinski, M. R. and Matthee, C. A. (2010) Remarkably low mtDNA control region diversity in an abundant demersal fish. *Molecular Phylogenetics and Evolution* 55: 1183-1188.

Von der Heyden, S. and Connell, A. (2012) Evidence of hybridisation within the genus Chrysoblephus and conserved nuclear sequences of South African sparids (Teleostei: Sparidae). *African Journal of Marine Science*, 34(4): 505-510.

Waples, R. S. (1998) Separating the wheat from the chaff: patterns of genetic differentiation in high gene flow species. *Journal of Heredity* 89: 438-450.

Ward, R. D., Zemlak, T. S., Innes, B. H., Last, P. R. and Hebert, P. D. (2005) DNA barcoding Australia's fish species. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 360(1462): 1847-1857.

Wares, J. P. (2002) Community genetics in the Northwestern Atlantic intertidal. *Molecular Ecology* 11(7): 1131-1144.

Warner, R. R. (1975) The adaptive significance of sequential hermaphroditism in animals. *American Naturalist*: 61-82.

Warner, R. R. (1988) Sex change in fishes: hypotheses, evidence, and objections. *Environmental Biology of Fishes*, *22*(2): 81-90.

Waters, J. M. and Roy, M. S. (2004) Phylogeography of a high-dispersal New Zealand sea-star: does upwelling block gene-flow? *Molecular Ecology*, 13(9): 2797-2806.

Watt-Pringle, P. A. (2009). Movement behaviour of three South African inshore sparid species in rocky intertidal and shallow subtidal habitats. MSc Thesis, Rhodes University

Weese, D. A., Fujita, Y., Hidaka, M. and Santos, S. R. (2012) The long and short of it: Genetic variation and population structure of the anchialineatyid shrimp Caridina rubella on Miyako-Jima, Japan. *Journal of Crustacean Biology*, *32*(1): 109-117.

Weinberg, J. R., Dahlgren, T. G., Trowbridge, N. and Halanych, K.M. (2003) Genetic differences within and between species of deep-sea crabs (Chaceon) from the North Atlantic Ocean. *The Biological Bulletin* 204(3): 318-326.

Weir, B. S. and Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution*, 1358-1370.

Western, D. (1979) Size, life history and ecology in mammals. *African Journal of Ecology* 17(4), 185-204.

Whitlock, M. C. and McCauley, D. E. (1999) Indirect measures of gene flow and migration: FST≠ 1/(4Nm+ 1). *Heredity*, 82(2): 117-125.

Wilkinson, T. A., Wiken, E., Creel, J. B., Hourigan, T. F. and Agardy, T. (2009) *Marine ecoregions of North America*. Instituto Nacional de Ecología.

Winkelmann, I., Campos, P. F., Strugnell, J., Cherel, Y., Smith, P. J., Kubodera, T., Allcock, L., Kampmann, M. L., Schroeder, H., Guerra, A. and Norman, M. (2013) Mitochondrial genome diversity and population structure of the giant squid Architeuthis: genetics sheds new light on one of the most enigmatic marine species. *Proceedings of the Royal Society of London B: Biological Sciences* 280 (1759): p.20130273.

Winkler, A. C. (2013) Taxonomy and life history of the zebra seabream, *Diplodus cervinus* (Perciformes: Sparidae), in southern Angola. MSc thesis, Rhodes University, South Africa.

Xia, X. and Xie, Z. (2001) DAMBE: software package for data analysis in molecular biology and evolution. Journal of heredity, 92(4): 371-373.

Zardoya, R., Castilho, R., Grande, C., Favre-Krey, L., Caetano, S., Marcato, S., Krey, G. and Patarnello, T. (2004) Differential population structuring of two closely related fish species, the mackerel (Scomber scombrus) and the chub mackerel (Scomber japonicus), in the Mediterranean Sea. *Molecular ecology* 13: 1785-98.

Zhang, A. B., Muster, C., Liang, H. B., Zhu, C. D., Crozier, R., Wan, P., Feng, J. and Ward, R. D. (2012) A fuzzy-set-theory-based approach to analyse species membership in DNA barcoding. *Molecular ecology*, *21*(8):.1848-1863.

Zink, R. M. and Barrowclough, G. F. (2008) Mitochondrial DNA under siege in avian phylogeography. *Molecular ecology*, 17(9): 2107-2121.

Appendix

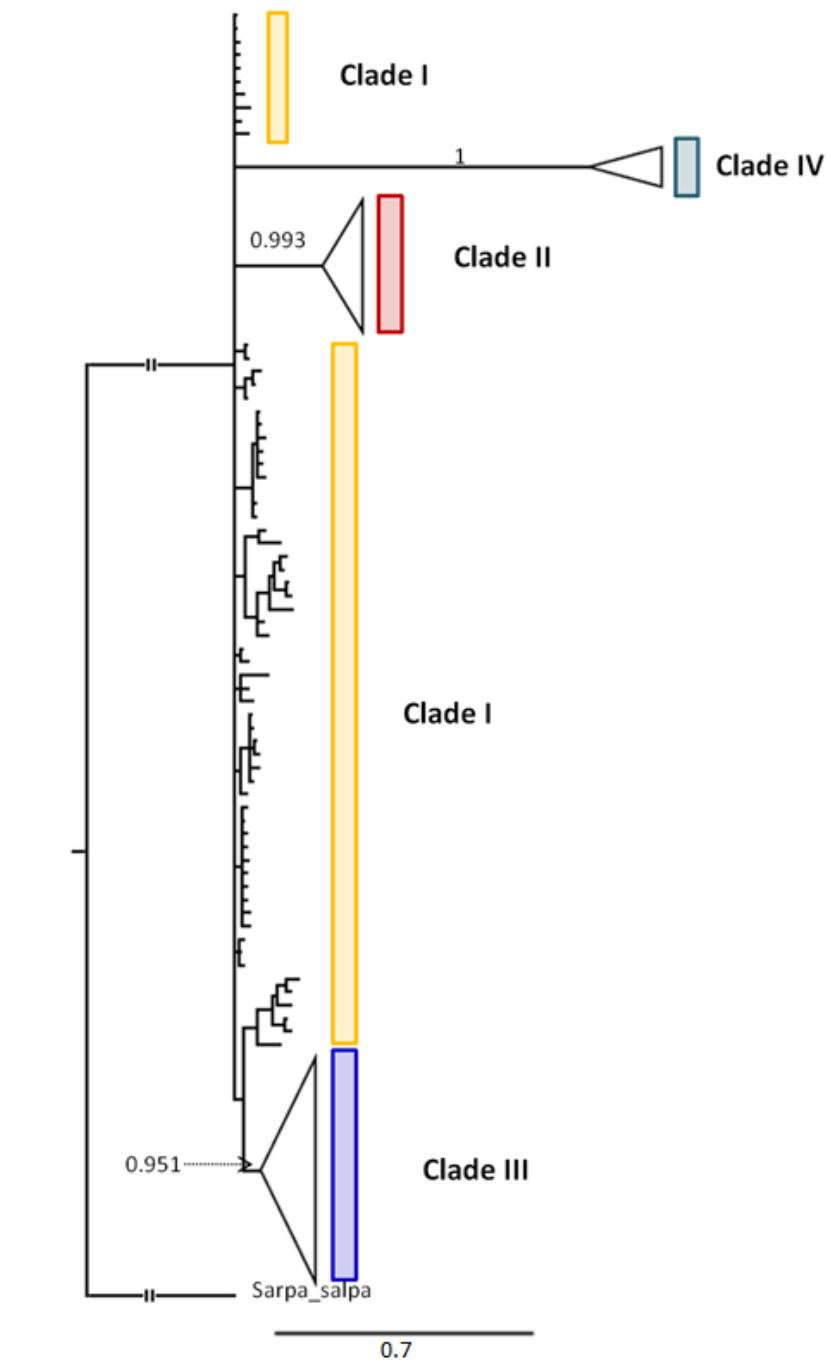

Figure A.1 Bayesian reconstruction of phylogenetic relationships among CR haplotypes of *L. mormyrus*. Identified clades (except clade I) have been collapsed to aid presentation and interpretation. Bayesian posterior probability support values are displayed on branches or are otherwise highlighted. Branches to outgroup have been cut (indicated by hashed lines) to aid presentation.

.

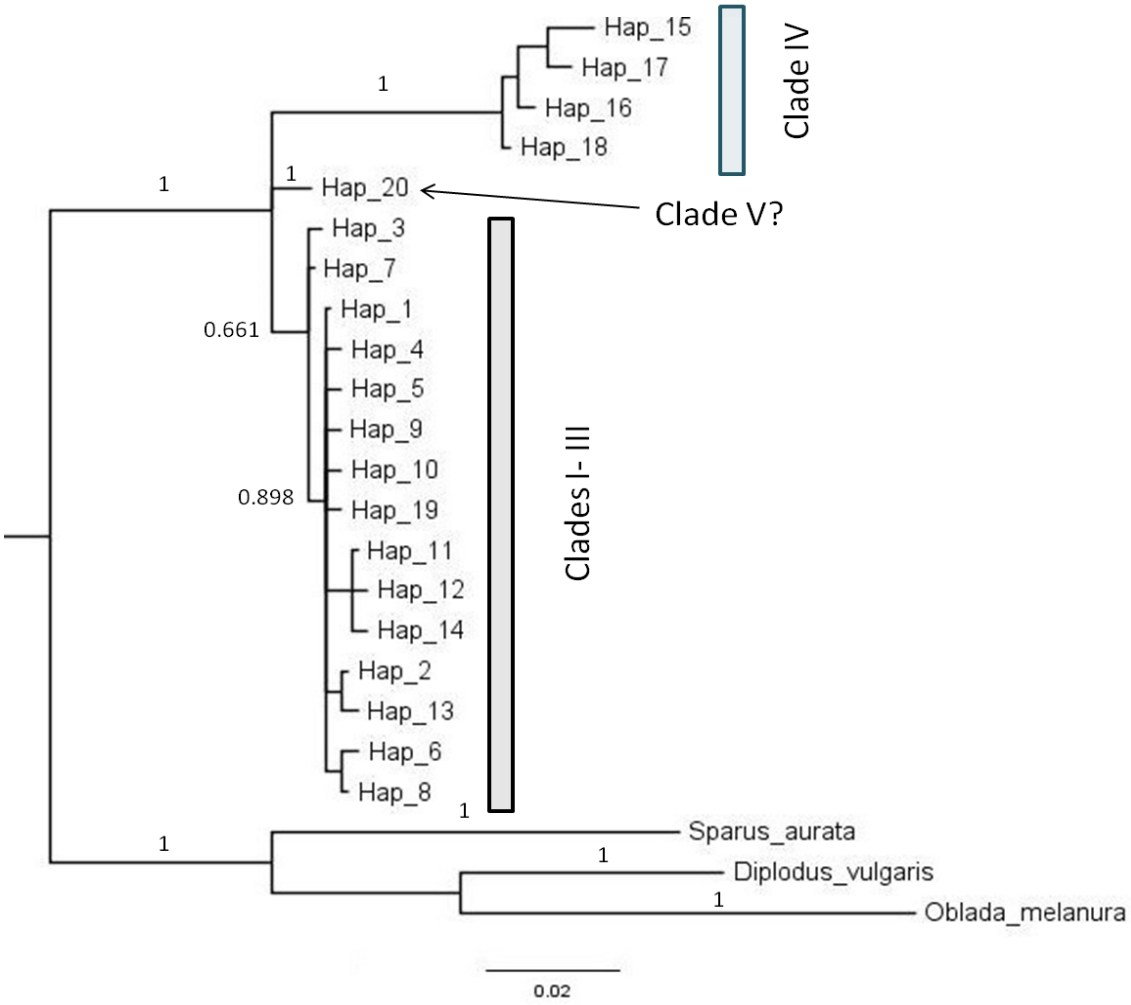

Figure A.2 Bayesian reconstruction of phylogenetic relationships among COI haplotypes of *L. mormyrus*. Bayesian posterior probability support values are displayed on branches or are otherwise highlighted.

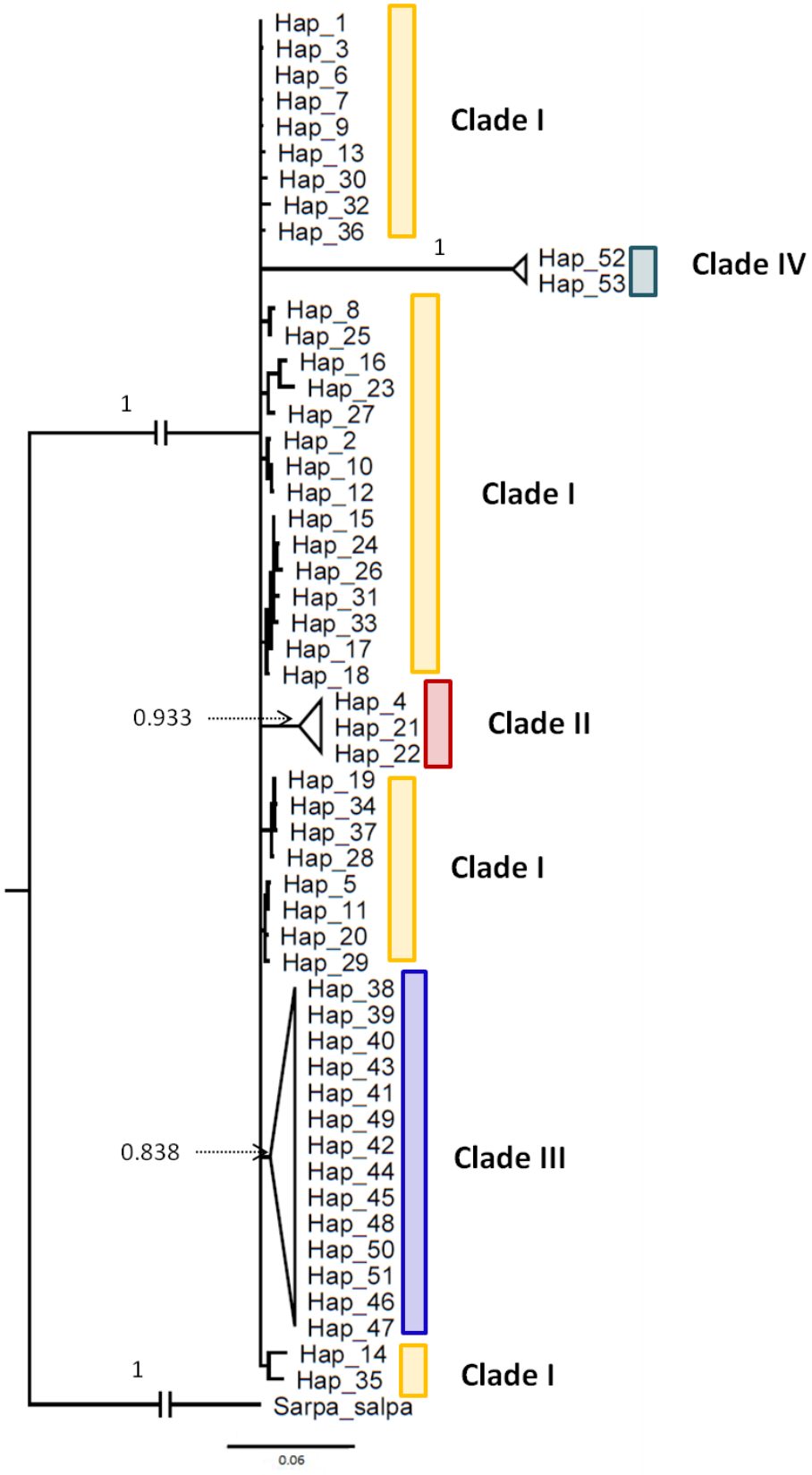

Figure A.3 Bayesian reconstruction of phylogenetic relationships among concatenated COICR haplotypes of *L. mormyrus*. Identified clades (except clade I) have been collapsed to aid presentation and interpretation. Bayesian posterior probability support values are displayed on branches or are otherwise highlighted. Branches to outgroup have been cut (indicated by hashed lines) to aid presentation.

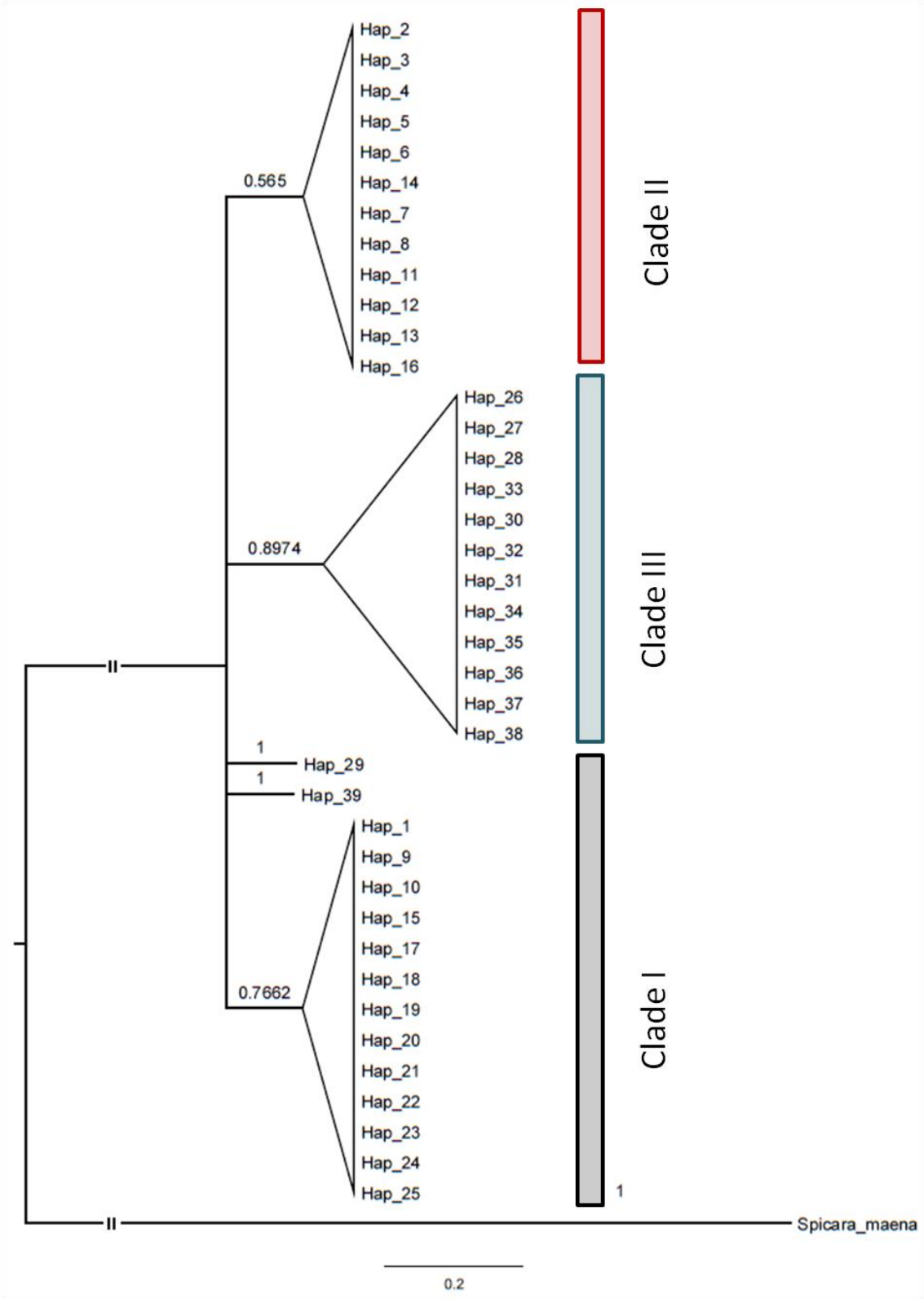

Figure. A.4 Bayesian reconstruction of phylogenetic relationships among CR haplotypes from Angolan, South African and Mediterranean *S. salpa*. Posterior probability support values are shown on the branches. Branches to outgroup have been cut (indicated by hashed lines) to aid presentation.

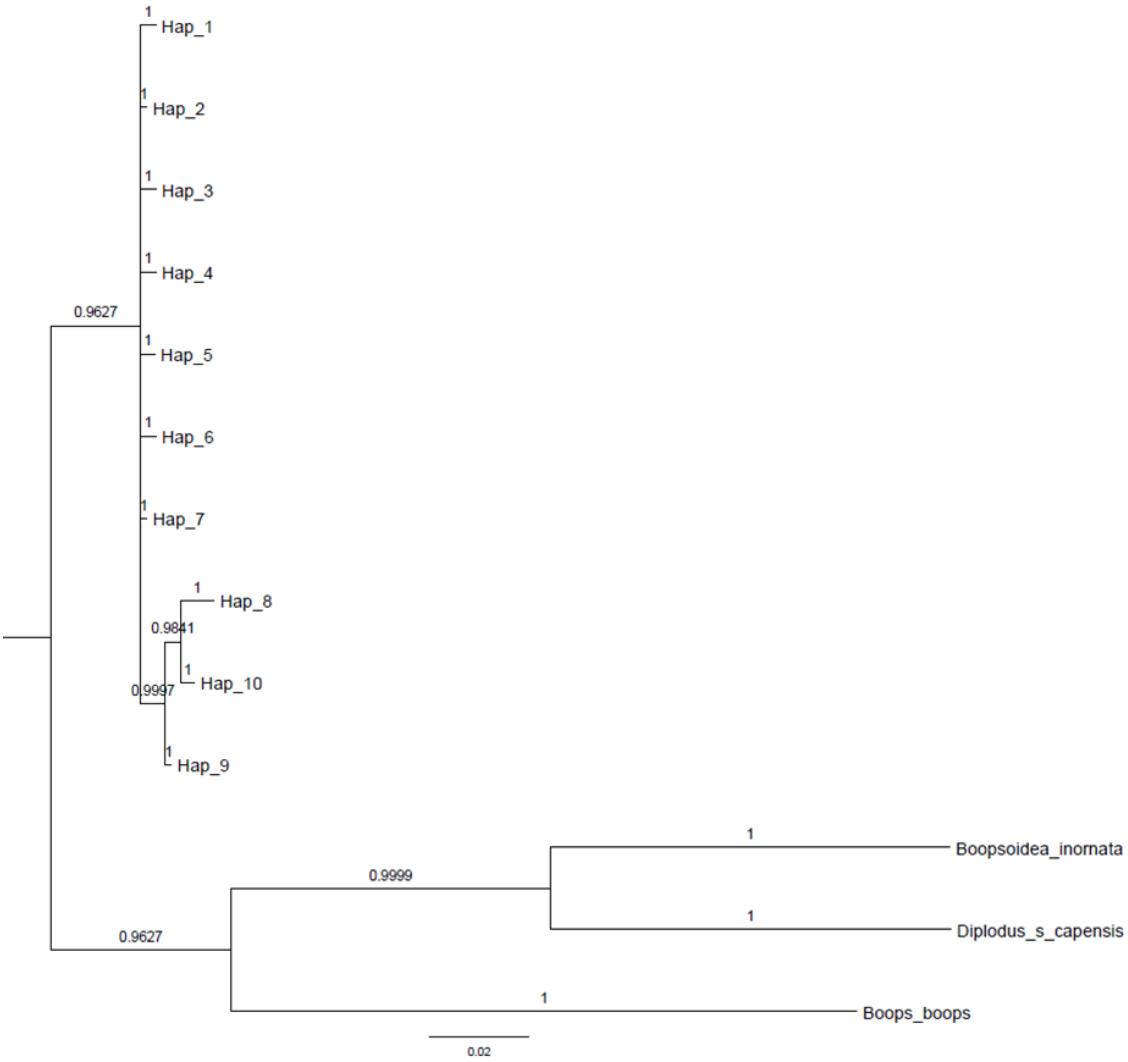

Figure. A.5 Bayesian reconstruction of phylogenetic relationships among COI haplotypes from Angolan, South African and Mediterranean *S. salpa*. Note that the Bayesian tree does not resolve any distinction between the Atlantic and Mediterranean *S. salpa*. Posterior probability support values are shown on the branches.

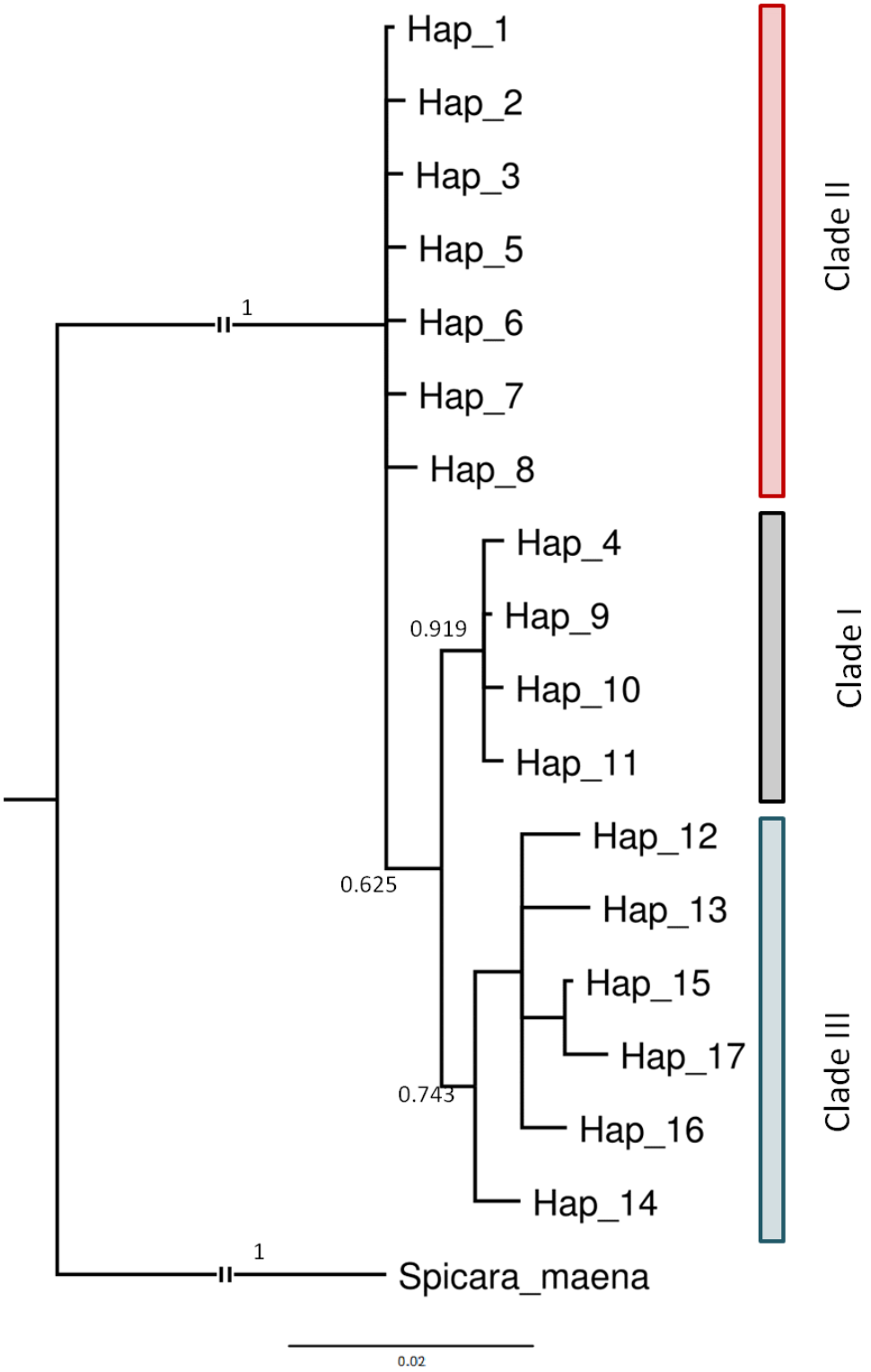

Figure. A.6 Bayesian reconstruction of phylogenetic relationships among concatenated COI CR haplotypes from Angolan, South African and Mediterranean *S. salpa*. Posterior probability support values are shown on the branches. Branches to outgroup have been cut (indicated by hashed lines) to aid display.

Table A.1 The seven microsatellite loci tested in the present study for all four study species, showing the Microsatellite ID, original species the primers were developed for, reference, forward and reverse primer sequences and the optimised Ta (°C) used in the present study for each species. LM= *L. mormyrus* and SS= *S. salpa*

ID	Original Species	Reference	Primers F	Primers R	LM Ta °C	SS Ta°C
DsaMS27	*Diplodus sargus*	Perez *et al*. (2008)	F: GCTCACTGTGCTGGCTCCACATCACC	R: GCGCTGTGCTTGCTGTCGGAGA	55	55
DsaMS34	*Diplodus sargus*	Perez *et al*. (2008)	F: AGATCAGATTTGCTGTGATAGCGTCCAAAG	R: ACTCCTGCAGCTCCTCCTGGGCTTC	55	55
Dvul33	*Diplodus vulgaris*	Roques *et al.* (2007a)	F: GCCGGGCTCGACATTGACACTGAA	R: GCAGCCAGCAGAGCTTAAAGAACT	-	50
Dvul38	*Diplodus vulgaris*	Roques *et al.* (2007a)	F: TCGGGCACAGATAGAAAGAAACAC	R: GAAGGAAGACGGATCTCAGGATGA	55	50
Dvul4	*Diplodus vulgaris*	Roques *et al.* (2007a)	F: GCGGTTATGTATACGTTGCGTTTA	R: TTGGCGTTGAACAGAAGTCAGACA	55	-
Dvul84	*Diplodus vulgaris*	Roques *et al.* (2007a)	F: GCTCGACGTGCACTCTGCCCTTGA	R: ATTCCCCAAATCCAGCACTCACAT	55	55
Omel58	*Oblada melanura*	Roques *et al.* (2007b)	F: GGCATTATTGTTCCATCATTACTCC	R: ATGGCATACAACCTGCATCAGAAG	-	55

Table A.2 Genetic diversity for *L. mormyrus* from mtDNA CR sequences. n: sample size; H: haplotype number; PH private haplotype number; *h*: haplotype diversity; π: nucleotide diversity. Sample ID codes can be found in Table 3.1.

	BEN	NAM	FLA	TOM	FAL	AGU
N	19	25	10	27	11	29
H	19	23	9	21	10	19
PH	13	19	4	14	4	13
***h* (SD)**	1.000 (0.017)	0.993 (0.013)	0.977 (0.054)	0.963 (0.026)	0.982 (0.046)	0.956 (0.023)
π(SD)	0.0304 (0.0157)	0.0215 (0.0111)	0.0119 (0.0068)	0.0270 (0.0138)	0.0163 (0.0091)	0.0095 (0.0052)

Table A.3 Genetic diversity for from *L. mormyrus* mtDNA COI sequences for individual samples. n: sample size; H: haplotype number; PH private haplotype number; π: nucleotide diversity; h: haplotype diversity. Sample ID codes can be found in Table 3.1.

	BEN	NAM	FLA	TOM	FAL	AGU
n	14	8	6	4	15	8
H	5	4	4	2	4	2
PH	1	2	0	0	2	0
h (SD)	0.725 (0.086)	0.750 (0.139)	0.800 (0.172)	0.500 (0.265)	0.6571 (0.0800)	0.536 (0.123)
π(SD)	0.0018 (0.0015)	0.0016 (0.0014)	0.0023 (0.0019)	0.0009 (0.0011)	0.0016 (0.0013)	0.0009 (0.0010)

Table A.4. Φ_{ST} values for the three identified CR Atlantic *L. mormyrus* clades. Significant values are denoted in **Bold.**

	Clade I	Clade II
Clade II	**0.695**	-
Clade III	**0.482**	**0.739**

Table A.5 Genetic diversity for *S. salpa* mtDNA CR sequences. n: sample size; H: haplotype number; PH private haplotype number; h: haplotype diversity; π: nucleotide diversity. Sample ID codes can be found in Table 3.2.

	LUC	FLA	REB	PEL
N	10	30	26	11
H	9	25	20	9
PH	7	23	18	7
h (SD)	0.976 (-0.054)	0.982 (-0.016)	0.969 (0.022)	0.964 (0.051)
π (SD)	0.0234 (-0.013)	0.02 (-0.0103)	0.0062 (0.0036)	0.0037 (0.0025)

Table A.6. Genetic diversity for *S. salpa* mtDNA COI sequences in individual sampling sites from Southern Africa. n: sample size; H: haplotype number; PH private haplotype number; h: haplotype diversity; π: nucleotide diversity. Sample ID codes can be found in Table 3.2. #

	FLA	INF	REB	PEL
N	24	4	8	8
H	3	1	2	2
PH	2	0	0	0
h (SD)	0.554 (0.053)	0	0.250 (0.180)	0.250 (0.180)
π (SD)	0.0010 (0.0009)	0	0.0004 (0.0006)	0.0004 (0.0006)

Table A.7 Pairwise Φ_{ST} values based upon the mtDNA CR sequence data between *S. salpa* CR clades. **Bold** indicates statistically significant Φ_{ST} values.

	Clade I	Clade II
Clade II	**0.774**	**-**
Clade III	**0.766**	**0.643**